Journal of Combinatorial Mathematics
and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

A Study on the Linguistic Adaptation of Language Modeling Techniques in Translating British Victorian Literature

Yanyan Lei[1,] ✉

[1] *College of Foreign language, Hechi University, Hechi, Guangxi, 546300, China*

ABSTRACT

This paper discusses the application of the neural machine translation model based on language modeling technology in British Victorian literature and its linguistic adaptation. Firstly, the linguistic features of Victorian literary works are analyzed, including thematic content and social background. Then the neural machine translation model based on language modeling technology is designed, and the text style migration method based on style representation is proposed to reproduce the linguistic features of the literary works. The performance of the translation models under the three fusion style methods is compared with five baseline systems, and the BLEU value, style migration accuracy, and style migration fluency of the machine translation model using the text migration decoding module are 37.49, 0.978, and 3.59, respectively, which are all higher than those of other models. Taking the translation of Wuthering Heights as an example, there is not much difference between this model and the human translation in terms of language adaptation evaluation. It shows that the machine translation model designed based on language modeling technology in this paper has better language adaptability for translating Victorian literature.

Keywords: language modeling technology, neural machine translation model, text style migration, Victorian literature

1. Introduction

Translation is a creative and artistic behavior activity. Literary translation is the process of translating novels, essays, poems and other literary works into another language. There are three things that are not easy to do in translation: one is to be faithful to the original, with fluent and understandable translation and elegant writing; the other is to conduct a comprehensive analysis and research of the original; and the third is to truly understand the original. Throughout China's translation career in recent years, great progress has been made, and its corresponding translation theories have also presented a hundred schools of thought and a hundred flowers blossoming [1-3]. Translation is a

✉ Corresponding author.
E-mail address: hcxywyx2024@126.com (Y. Lei).
Received 20 February 2024; Revised 05 April 2024; Accepted 10 November 2024; Published Online 15 April 2025.
DOI: 10.61091/jcmcc127a-183

science and an art. Translators should flexibly utilize the existing skills, and they should be good at creating new skills in order to make the translated text convey the meaning and achieve the purpose, and to be both in shape and spirit. In Fu Lei's words, "Although the translation is close to the tongue, it should be based on artistic cultivation. Without the beauty of the mind, without warm sympathy, without proper appreciation, without considerable social experience, without sufficient common sense, it is difficult to completely understand, that is, or to understand, may not be able to deeply comprehend." Therefore, the translator has to understand and appreciate the original work first, and then needs to further express what he understands and appreciates, faithfully and movingly [4-6]. In fact, this expression is a process of re-creation, which requires the translator to go deep into the spiritual field of the original work to explore first, including the understanding of the author of the original work, as well as the background of its creation, the purpose of its creation, the style and tone of its writing, and the translation should be as comprehensive as possible to have all the literary information of the original work. As a literary re-creation, it means that the translation should try its best to play within the limited space, and it can achieve the goal of immersing the readers of the country into the exotic atmosphere depicted by the original author. In addition the translation should have a high degree of linguistic adaptability, it should try to conform to the laws of the national language and read in accordance with the habits of the national readers, but it also retains at the same time a strong exotic flavor [7-9].

With the continuous progress of computer technology and language modeling technology, translation is moving from manual translation to machine translation in a comprehensive way. Machine translation is the use of computers to convert one natural language (source language) into another natural language (target language), and has gone through several major stages, such as dictionary-based translation, rule-based translation, statistical-based translation, and machine-learning-based translation [10-12]. In the past two years, with the continuous progress of algorithms and arithmetic power, large language models based on massive, high-quality training texts have begun to appear, and their parameter sizes can easily reach hundreds of millions or even tens or hundreds of billions. Big language models have cross-domain and cross-task generalization, and have excellent performance on several benchmark tests, thus, they are also known as cornerstone models or foundation models [13-15]. The Big Language Model has strong language comprehension and language generation capabilities, and is able to perform various tasks including translation of literary works according to instructions. In addition, big language models also have emergent capabilities such as situational learning and chain of thought, and are able to optimize and adjust the prediction results generated by themselves using additional information in the context, which provides the possibility of machine translation paradigm innovation. The emergence of big language modeling has brought machine translation into a new era, on the one hand, big language modeling can be used directly as a machine translation tool, on the other hand, big language modeling can also be used as an important auxiliary tool for translation teaching to improve the efficiency and effect of translation teaching [16-20].

In this paper, a neural machine translation model based on character-level language modeling is designed, and the model combines character-level coding techniques with denoising self-encoding-based language modeling for the translation of Victorian literature. And a text style migration method based on style representation is incorporated into the model to make the model linguistically adaptable. The study proposes three fusion style representation methods at this stage: embedded representation pre-training model, style context fusion module, and text migration decoding module. The performance of the models under different fusion methods is explored, and the best-performing combined model is selected for the translation of Wuthering Heights literary works and compared with the manual translation, and the linguistic adaptability of the model is analyzed according to the evaluation results.

2. Linguistic Features of British Victorian Literature

The Victorian era refers to the years of Queen Victoria's reign in England (1837-1901), and Victorian literature is British literature during the reign of Queen Victoria. The most notable feature of British Victorian literature was the rise of the novel. Famous novelists of this period include Charles Dickens, William Thackeray, the Three Brontë Sisters, George Eliot, and Thomas Hardy. Victorian literature has a significant international influence as the literature of the era with rich connotations and profound influences.

2.1. Thematic content

The struggle of the individual to adapt to society was a prominent theme in Victorian fiction. By the late nineteenth century, the view in Thomas Hardy's novels became increasingly pessimistic, with Hardy repeatedly reminding, alerting individuals to the range of conflicts in which they are involved: city and country, upper and lower classes, men and women, faith and disorientation. Hardy's novels were all published in the wake of Darwin's theory of evolution in The Origin of Species (1859), and his work repeatedly reflects this vision: the Victorian populace realizing that they were living in a state of anxiety provoked by a Darwinian world that questioned God's creation. As a result, his series of hard-to-please individuals seem to be held hostage by the complex and contradictory uncertainty and fatal existentialism that this anxiety provokes.

2.2. Social context

The Victorian era centered around the political career of Queen Victoria. She was crowned in 1837 and died in 1901. Many changes took place during this period - brought about by the Industrial Revolution. It is therefore not surprising that literature from this period often focuses on social reform. Indeed, Victorian literature presents a diversity of changes that stemmed precisely from the changing context and ideological shifts in Victorian society. 1859 saw the publication of Charles Darwin's book Natural Selection in the Origin of Species, and historians, philosophers, and scientists alike began to apply the ideas of evolution to the new field of study of the human experience. Traditional ideas about the nature of man and his place in the world were shaken to their core.

3. Language modeling techniques and their application in translation

This paper proposes a neural machine translation model based on character-level language modeling for the translation of Victorian literature. It also incorporates a text style migration method based on style representation to make the model linguistically adaptable.

3.1. Character-level language modeling and overall model architecture

The language of Victorian literature belongs to the 19th century English, which has some differences in vocabulary, grammar and spelling with modern English. Aiming at the problems such as the scarcity of parallel corpus in the language of Victorian literary works, this paper proposes a neural machine translation model based on character-level language modeling, which combines the character-level coding technology with the denoising self-encoding-based language modeling, and greatly improves the performance of neural machine translation. The overall model consists of a Transformer encoder submodule and a decoder submodule in the form of an encoder-decoder pairwise framework. Character-level encoding input is used in the encoder side of the model, and word-level decoding output is still used in the decoder side. The overall model framework is shown in Figure 1.

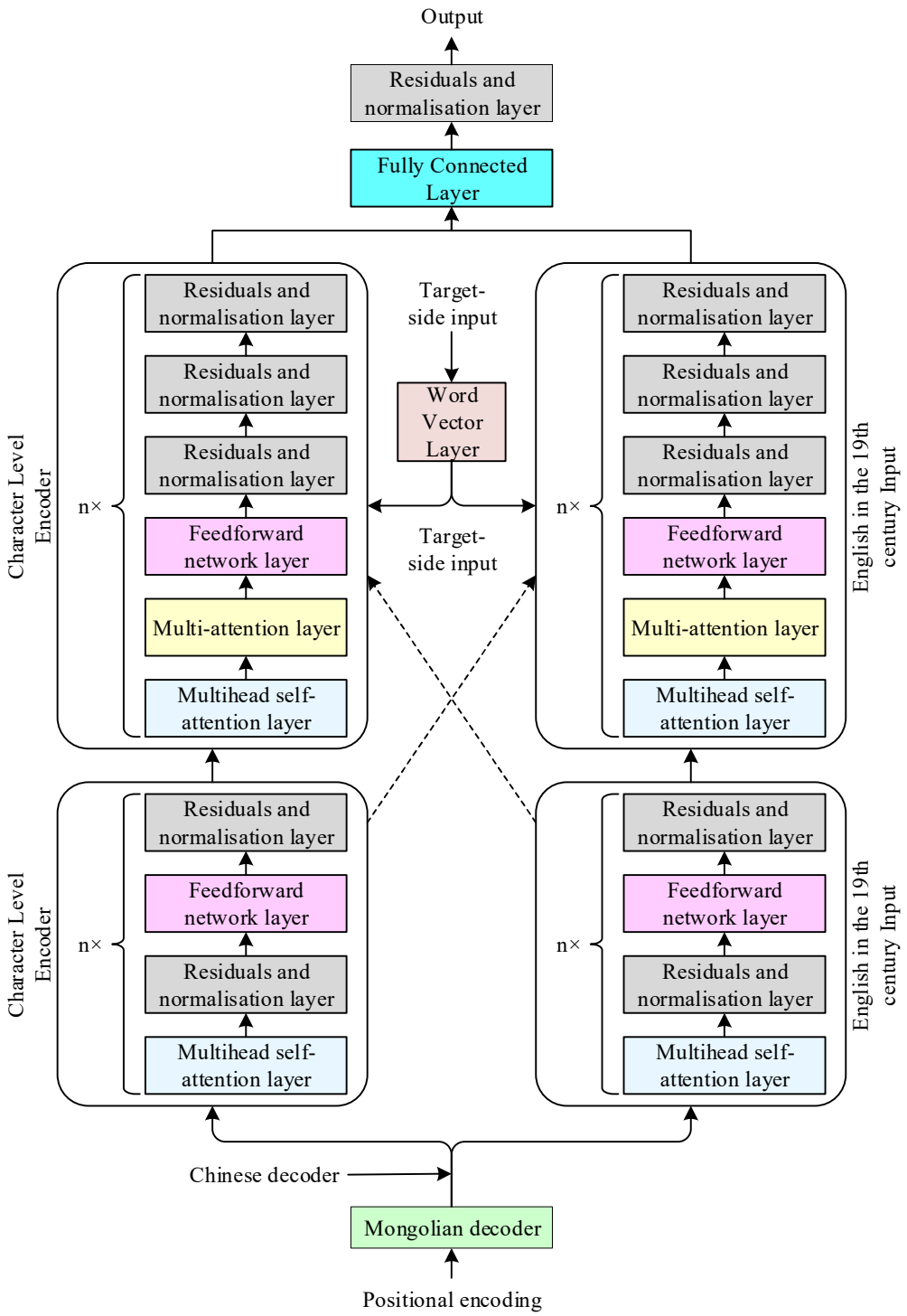

Fig. 1. Machine translation model based on character level language modeling

The cross section of the proposed character-level language modeling-based neural machine translation model represents the character-level language modeling process. The overall architecture includes character-level encoder, bilingual encoder, bilingual decoder, full connectivity layer, residual connectivity and normalization layer. In the training phase, the decoder receives the output from the encoder will be combined with the utterance input from the target to translate the output, and the loss rate is obtained to update the model parameters for model training iterations [21].

The character-level language coding method proposed in this model is based on the denoising self-coding technique, which firstly requires the input sentences to be segmented by adding the separator

"{}" outside each word, secondly, the identifier "<eos>" needs to be added at the end of the sentence to separate the sentences, and finally Noise processing is added to the bilingual-based monolingual input data. Three kinds of noise are used in the paper: firstly, some phrases in the sentence are deleted with a certain probability. The second is to randomly select a fixed length window and disrupt the words within the window. The third is to select a word with a probability of 10% and then replace it with the character "<mask>" with a probability of 80%, leaving the phrase unchanged with a probability of 20%. This processing has a certain probability of randomness, and the processed samples will not be too homogeneous, which is convenient for the model to learn and process semantic features.

The Chinese and 19th century English subword-level monolingual corpus is trained, the processed characters are composed into a dictionary, the vectors are extracted and matrixed, the positional encoding is added, and then the Chinese or 19th century English encoder-decoder is operated, and then the sentence before noise is obtained through the fully-connected layer and the residuals and normalization layer, so that the machine can be trained to understand the construction of the utterance and the grammatical meaning. Let the Chinese monolingual corpus be z, and the sentence after the participle processing of the Chinese monolingual corpus be x, and the sentence after the noise processing of x be obtained as $C(x)$. Let the 19th century English monolingual corpus be m, and the 19th century English monolingual corpus be the participle location coding and the 19th century English denoising self-coding process, and the objective function of this character-level modeling model is:

$$L^{zm} = E_{x\sim z}\left[-\lg P_{z\to z}\left(x\middle|C\left(x\right)\right)\right] + E_{y\sim m}\left[-\lg P_{m\to m}\left(y\middle|C\left(y\right)\right)\right] \tag{1}$$

The model can realize source-target language and target-source language translation. Set s comes from the source monolingual corpus z, which is subjected to segmentation and character encoding operations, and after being encoded by the source language encoder, it enters into the target language decoder to be decoded, and then outputs through the fully connected layer and the residuals and normalization layer, obtaining the translated target language $g(s)$. Set t comes from the target monolingual corpus m, which is generated by a translation model, and the target language $g(t)$ is generated. The pseudo-parallel pairs generated by the character-level neural network-based translation model can be used to iteratively train the translation model, and the corresponding minimum loss function is:

$$L^{loss} = E_{t\sim m}\left[-\lg P_{z\to m}\left(t\middle|g\left(t\right)\right)\right] + E_{s\to z}\left[-\lg P_{m\to z}\left(s\middle|g\left(s\right)\right)\right] \tag{2}$$

The overall model consists of language modeling training and translation training, with iterative training with a minimization goal:

$$L = L^{zm} + L^{loss} \tag{3}$$

L^{zm} and L^{loss} represent the loss rate for language modeling training and translation training, respectively, which decreases as the model training iterates, and the goal of each round of model training is to pursue the value of minimizing L.

3.2. Text style migration method based on style representation

3.2.1. Stylistic Representation Model Based on Label Embedding and Graph Networks. The purpose of this part of the experiment is to investigate how to represent the textual style features of Victorian literature. Combined with the label embedding technique, this paper embeds the style labels in the style dataset of Victorian literary works to represent them, and utilizes the graph network method to improve the representation capability of the model, and the resulting style label embedding is regarded as the style representation.

Common graph embedding algorithms can be categorized into wandering-based algorithms and neural network-based methods. The typical wandering-based algorithm is Deep Walk, while the deep network-based algorithm is represented by graph neural network (GCN). In this paper, we combine the label embedding with graph network, and utilize wandering-based graph embedding algorithm to embed the representation of style label nodes and word nodes, and regard the embedding vectors of the style labels as the style representation.

1) Deep wandering based graph node sampling. The Node2Vec algorithm introduces control parameters P and Q compared to the traditional DeepWalk algorithm, and its sampling process is shown in Fig. 2. During the random walk, the algorithm has sampled node t and node v, and is currently staying at node v, denoted as (t,v). The algorithm will determine whether to walk to the next node x. From equation (4), when node t is the same as node x, the probability of node x being sampled is 1, and when node t is connected to node x, the probability of node x being sampled is $1/q$. If node t is not connected to node x, the probability that node x is sampled is $1/p$:

$$\alpha_{pq}(t,x)=\begin{cases}1/p & ifd_{tx}=0\\ 1 & ifd_{tx}=1\\ 1/q & ifd_{tx}=2\end{cases} \tag{4}$$

For the adjustment of parameters P and Q, it is the main method to optimize the effect of Node2vec algorithm. For the return probability parameter p, when $p>\max(q,1)$, the sampling algorithm backtracks with lower probability, and when $p<\max(q,1)$, the sampling algorithm backtracks with higher probability, resulting in aggregation of sampled nodes. For the access parameter q, when $q>1$, the sampling algorithm reflects out of the characteristics of width-first search, and when $q<1$, the algorithm reflects the characteristics of depth-first search. Set the number of samples is N, the maximum depth of sampling is D, then we can get the collection of sampled nodes K, after getting the sampling collection, we can input it to the node embedding algorithm for pre-training, and get the embedded representation of all the nodes in the graph.

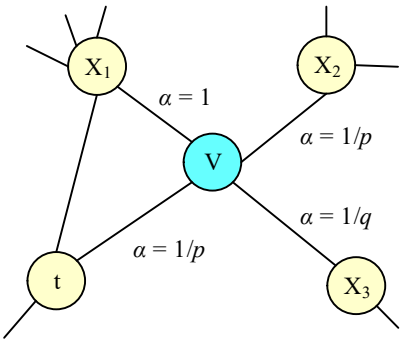

Fig. 2. Node2vec algorithm

2) Skip Gram based node embedding algorithm. Skip Gram is one of the training paradigms for Word2Vec model, which is commonly used for text pre-training in natural language processing to obtain word representations. In the Node2Vec algorithm, the node set K is obtained through the wandering sampling algorithm, which contains a total of n nodes, and for each node in the set, the embedding dimension is defined to be d, which yields the embedding matrix $E_{n\times d}$. Skip Gram first constructs a neural network based on the number of nodes in the training data and utilizes a normally distributed random initial representation matrix to represent the word nodes and style nodes based on the embedding matrix. After representing the word nodes and style label nodes based on the embedding matrix, the model prediction results can be obtained after going through the fully

connected layer and the downstream classification network, from which the complete SkipGram model can be constructed:

$$\max \sum_{u \in V} \log P\left(N_s(u) \middle| f(u)\right) \tag{5}$$

$$P\left(w_i\right) = f\left(w_i\right)^{3/4} \Big/ \sum_{j=0}^{n} \left(f\left(w_j\right)^{3/4}\right) \tag{6}$$

The training process of Skip Gram model is shown in the above equation, Skip Gram model adopts the method of predicting the surrounding words from the center word, for the style statement $W = \{w_{n-2}, w_{n-1}, w, w_{n+1}, w_{n+2}\}$, i.e., predicting w_{n-1}, w_{n+1}, etc., from W_n, and its predicted target is shown in Equation (5), and the positive samples of its prediction are the target words, however, due to the excessive negative samples, the algorithm adopts the negative sampling technique in Equation (6) at the same time, which reduces the number of negative instances in the computation of the loss , which reduces the amount of computation in the training phase of the model.

3.2.2. Text style migration methods. The text style migration neural network model based on style representation is shown in Fig. 3. According to the style representation model based on label embedding and graph network proposed above, the static embedding representations of word nodes and style label nodes can be obtained under the same vector space by constructing local and global fusion of text graphs, modeling co-occurring information in the same utterance and across utterances, and using the Node2Vec algorithm to perform a deep wandering. In the text generation stage, the model initializes the text embedding matrix using the learned static word representation, and the encoding model models the text through the Transformer model to obtain a high-dimensional semantic representation of the style text of Victorian literature. Meanwhile the model defines the style embedding matrix, and the style embedding representation can be directly obtained by specifying the style. In the contextual style fusion module, the contextual information is fused with the style representation information at the decoding end, and the textual representation of opposing styles is obtained through content reconstruction and style migration.

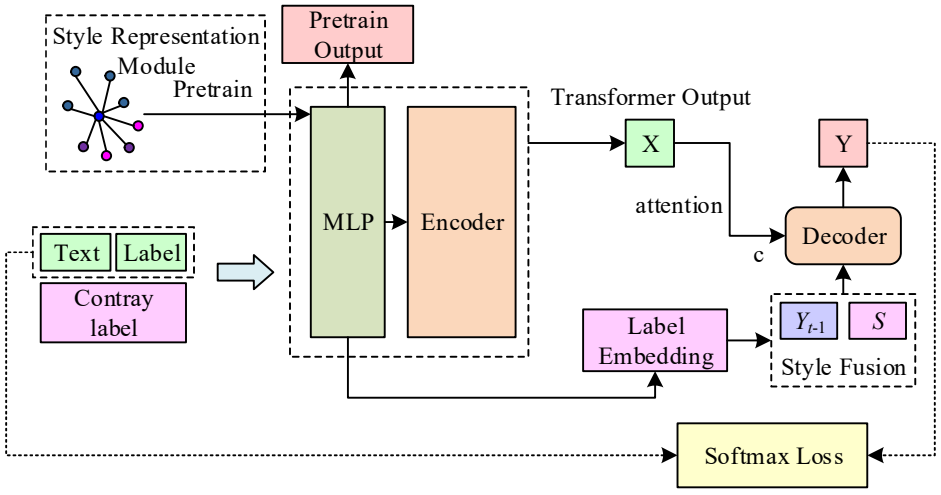

Fig. 3. Text style transfer method based on style presentation

1) Text style migration dataset. This paper collects 25 collections of works from the Victorian era in Britain, totaling 633 works as the initial dataset. Statistical analysis is carried out on this dataset to analyze the style of Victorian literary works from the perspectives of words and sentences, and to conclude that the Victorian era was more conventional in terms of words and sentences. Then, the

method of back-turning was adopted, Baidu Translation was retrieved for 19th century English to Chinese translation, and the modern text with the Victorian style was obtained, and it was combined with the original Victorian text into sentence pairs one by one, and the parallel corpus of "modern text - Victorian style text" was constructed as a training dataset. The data is then subjected to preprocessing steps such as word splitting, symbol normalization, tokenizer, BPE coding, data filtering, and dataset partitioning.

2) Text style migration based on style representation. The purpose of this part of the experiment is to study how to combine the style representation methods with the text generation model and further accomplish the text style migration task. In this paper, we try three methods of fusing style representation and construct a text style migration model based on the sentence representation and migration generation module that fuses style information.

(1) Pre-training model for embedded representation (CW). In this paper, we utilize the style representation method based on label embedding proposed above to obtain word representation and style embedding representation. The text graph based on local and global fusion is constructed, sampled using the wandering algorithm Node2Vec, and pre-trained using SkipGram to obtain the static embedding representation of words and style labels under the same vector space [22].

(2) Style context fusion module (CS). Based on the style representation method, the word representation matrix $E_{n\times d}$ of the stylized text can be obtained, where n is the number of all words in the text, and d is the dimension of the word embedding. At the same time, we can obtain the style embedding representation $S_{1\times d}$, which can be operated with the sequence vector after stacking the style representation vectors, and d is the dimension of the style label embedding, and the dimensions of the two are equal:

$$C = C + wS \tag{7}$$

$$C = w * concat(C, S) + b \tag{8}$$

$$P_t = w * concat\left(Emb\left(Y_{t-1}\right), S\right) + b \tag{9}$$

The input text is encoded by Transformer to obtain its sentence sequence representation vector $C_{m\times d}$. In the decoding stage, let the output word of the previous time step be Y_{t-1} and the decoded input be P_t. In this paper, we try several ways to fuse the contextual representation with the stylistic representation. The first way, as shown in equation (7), uses direct weighting to add the stacked style representation with the sentence sequence vector and adjusts the degree of fusion by weights. The second way is shown in equation (8), where the stacked style representation is spliced with sentence sequence vectors and then fused using fully connected networks. The third way is shown in equation (9), where the inputs of each time step in the decoding prediction stage are spliced with the style vectors and the fusion vectors are obtained by the fully connected network, and the prediction target is obtained by decoding.

(3) Text migration decoding module (DS). In this paper, the Transformer-based decoder is utilized to reconstruct the text content, and through the style fusion module, the style representation is fused with the text generation process. Let the input style statement $I = \{w_0, w_1 \ldots w_{n-1}\}$, then its corresponding prediction target is $\{w_1, w_2 \ldots w_n\}$. It can be seen that the target text in the training stage is equivalent to the result of the overall left shift of the source text, and the model ensures the thematic consistency between the generated text and the input text by reconstructing the source text:

$$Y_t = Decode\left(Y_{t-1}, S, C\right) \tag{10}$$

$$o_t = soft\max\left(Y_t, t\arg et\right) \tag{11}$$

The structure of the model is shown in Fig. 4. As shown in Eq. (10), the Transformer Decoder, in the decoding stage, obtains the current time-step output based on the target style representation, the

output of the previous time-step and the encoded output, and further calculates the loss function based on Eq. (11) by using Y_t and the target value, which serves as the optimization objective of the decoding model. Using methods such as Teacher Forcing, the model is updated for learning, and the style representation is integrated into the decoding process under the premise of ensuring content consistency, making the translation of Victorian literature linguistically adaptable.

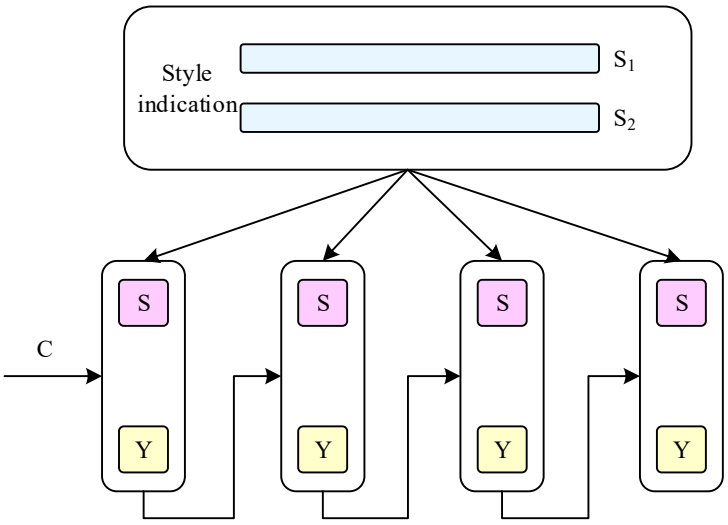

Fig. 4. Decoding module combined with style representation

3.3. Assessment of indicators:

1) BLEU score. In the supervised dataset, the model conversion ability can be measured by calculating the similarity between the model generated sentences and the target sentences. BLEU is a commonly used evaluation method for automatically evaluating the merits of translation results, which compares and counts the number of co-occurring n-tuples (N-grams) in the reference sentence and the generated sentence, and then divides the number of co-occurring n-tuples by the number of all the n-tuples in the reference translation, to get the evaluation In this paper, the BLEU score is used to measure the accuracy of the conversion result: the higher the BLEU between the generated sentence and the target sentence, the better the conversion accuracy is [23].The BLEU is calculated as follows:

$$BLEU = BP \times \exp\left(\sum_{n=1}^{N} w_n \log p_n\right) \tag{12}$$

where denotes the weight of the co-occurring n-tuple, is the proportion of co-occurring n-tuples, N is the length of the longest n-tuple (usually taken as 4), and BP is the penalization factor, used to penalize the generated sentences whose lengths are less than the reference sentence.

2) Style migration accuracy. Generally use the text classifier to complete the style migration accuracy calculation, first based on the source style and target style data to train the binary classification model, the model structure can be used RNN or pre-training model BERT. To get the machine translation, and then use the classification model for scoring, according to the score to determine whether the sentence expression style is the target style. The calculation formula is shown in (13):

$$l_{style} = \begin{cases} style1(positive) & output \le 0.5 \\ style2(negative) & output > 0.5 \end{cases} \tag{13}$$

Where OUTPUT is the score of the normalized output of the text classifier.

3) Style migration fluency. The purpose of language modeling is to describe the probability of occurrence of a text sequence can effectively assess the fluency of a sentence. Given a sentence $x=(x_1,\ldots,x_m)$, its probability is calculated as shown in (14):

$$p(x)=p(x_1)\,p(x_2|x_1)\,p(x_3|x_1,x_2)\ldots p(x_m|x_1,\ldots,x_{m-1}) \tag{14}$$

It is difficult to compute $p(x)$ directly, and such computation leads to high parameters and complexity of the model. The problem is simplified by introducing Markov's assumption, which assumes that each word in the text sequence is only related to certain words before it, reducing the difficulty of computing $p(x)$. The n-grams language model is to consider the previous $n-1$ historical words to generate the current word, and these probability parameters are obtained by calculating from large-scale text data. The formula for the n-grams language model after introducing Markov's assumption is shown in (15):

$$\begin{aligned} P(x) &= P(x_1,x_2,x_3,\ldots,x_m) \\ &= P(x_1)P(x_2|x_1)P(x_3|x_1,x_2)\ldots P(x_m|x_1,x_2,\ldots,x_{m-1}) \\ &\approx P(x_1)P(x_2|x_1)P(x_3|x_2,x_1)\ldots P(x_m|x_{m-n+1},\ldots,x_{m-1}) \end{aligned} \tag{15}$$

4. Language adaptation analysis

4.1. Model evaluation

4.1.1. Baseline systems. The experiments compare the neural machine translation model based on character-level language modeling proposed in this paper with the following baseline systems under three fused stylistic representations.

1) Base-Model: this method uses a generic translation model on a stylized test set.

2) Pivot-Rule: This method first automatically converts the language of the input text using the translation model, and then rewrites the previous text to convert it to the target stylized sentence by using some custom rules (e.g., case-switching, abbreviation, repeated punctuation, etc.).

3) Pivot-Model: This method first uses the translation model to automatically convert the language of the input text, and then uses the text style migration model to convert the source style sentences to the target style sentences.

4) Teacher-Student: This method uses the text style migration model as the teacher model and the translation style migration model as the student model, and the student model is trained on the pseudo-parallel corpus generated by the teacher model.

5) Back-Translation: this method trains two reverse-direction models, which are the text style migration model from target style to source style and the translation model from target language to source language, and then utilizes the above two models to expand the corpus size by back-translating large-scale target style monolingual data.

4.1.2. Experimental results. The experimental results on the previously constructed dataset are shown in Table 1.The performance of Base-Model is poor, and the translation results for 19th century English are skewed towards the modern style, which has a great stylistic difference from Victorian works, so it is necessary to train a style-specific translation model to improve the quality of translations. With the introduction of a text style migration model, the BLEU scores of the pivot-based methods (Pivot-Rule and Pivot-Model) improved significantly, but the translation style migration accuracy was relatively low, indicating that multiple models decoding in a pipelined fashion introduce noise and create error transmission problems. In this paper, the style representation fusion method (Our+DS) using decoding splicing improves 12.74 BLEU points compared to the Back-Translation method, achieving the best current results. It is shown that the use of the text migration decoding

module can lead to optimal style migration results for neural machine translation models based on character-level language modeling.

Table 1. Translation style migration results in data sets

Method	Style migration accuracy	Style migration fluency	BLEU
Base-Model	0.041	2.71	8.33
Pivot-Rule	0.866	3.16	19.97
Pivot-Model	0.871	3.31	21.9
Teacher-Student	0.883	3.47	22.67
Back-Translation	0.899	3.47	24.75
Our+CW	0.938	3.47	26.7
Our+CS	0.943	3.56	27.39
Our+DS	0.978	3.59	37.49

In order to evaluate the effect of sentence length on the model's linguistic adaptation, this experiment calculates BLEU scores for the translation results of different sentence lengths in the test set, and the experimental results are shown in Fig. 5. The neural machine translation model based on character-level language modeling using the text migration decoding module has the best performance on most sentence lengths. When the length of the source sentence increases, the BLEU scores of the translation results obtained by all methods decrease to some extent, and the longer the sentence, the higher the probability that the model loses information during translation. When the sentence length is small, the encoding ability of each model can obtain good translation results, for example, when the sentence length is between 5 and 10, the performance of the methods in this paper is at a high level except Back-Translation. After the sentence length exceeds 30, the performance difference between different methods is obvious, and the character-level language modeling-based neural machine translation model using the text migration decoding module improves the performance of Back-Translation by 25.2 BLEU points over Back-Translation's method.

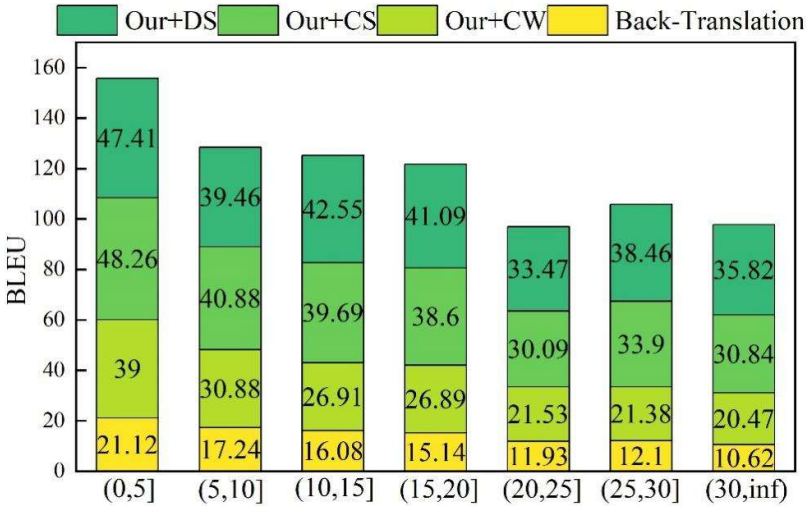

Fig. 5. The influence of sentence length

4.2. Simulation Tests

In order to further test and validate the neural machine translation model based on character-level language modeling using the text migration decoding module, the processing of syntax, semantics, and utterance ambiguity is analyzed as the basis for evaluating the translation results. Using the text style migration dataset constructed in the previous section, a five-point Likert scale is adopted, and the

translation is evaluated and scored by 30 experts, with the scores ranging from “strongly agree” to “strongly disagree” set at “5-1”.

The syntactic evaluation of the translations was analyzed in terms of grammatical correctness, sentence structure, and dependency. The evaluation results are shown in Figure 6. The number of experts who strongly agree with the three evaluation indexes is above 40%, which indicates that the syntactic treatment of the translated text under the model of this paper is ideal.

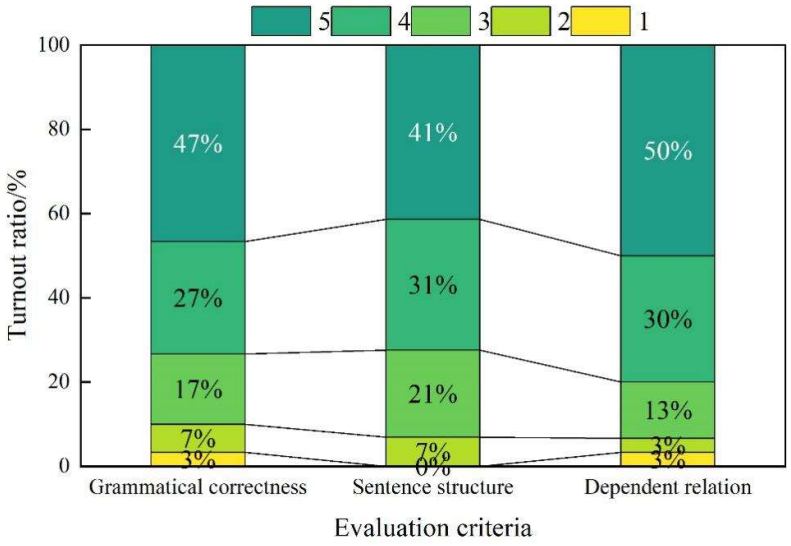

Fig. 6. The evaluation of the syntax

The evaluation of the semantics of the translation was analyzed in terms of meaning conveyance, fidelity and fluency. The evaluation results are shown in Figure 7. The total number of people with ratings of “4” and “5” for the three indicators is 67%, 66% and 70% respectively, and the number of people who are satisfied is high.

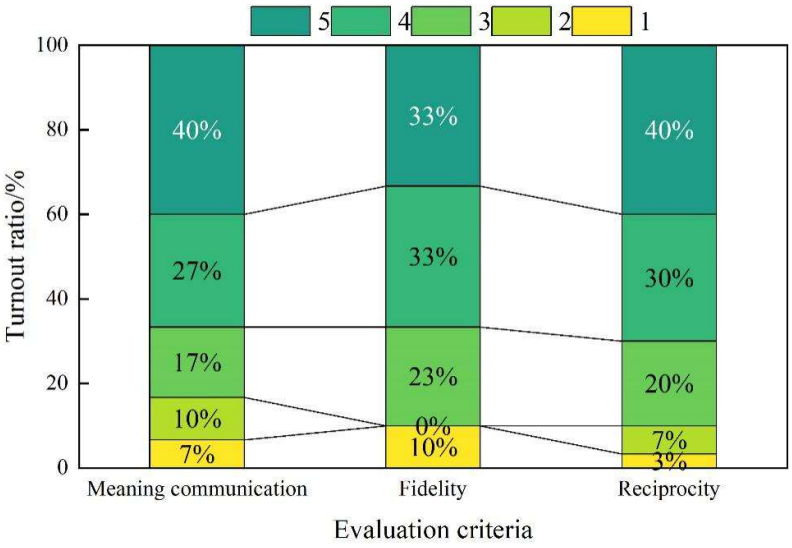

Fig. 7. The evaluation results of the semantic evaluation

The processing of utterance ambiguity was judged in terms of the model processing of lexical, grammatical, referential, modification, and tense aspects, respectively. The evaluation results are shown in Figure 8. The model's treatment of lexical ambiguity is the most effective, and the number of people evaluated with a score of 5 reaches 47%. It shows that although the 19th century English

vocabulary samples are sparse, the model in this paper can still handle lexical ambiguity better after improvement.

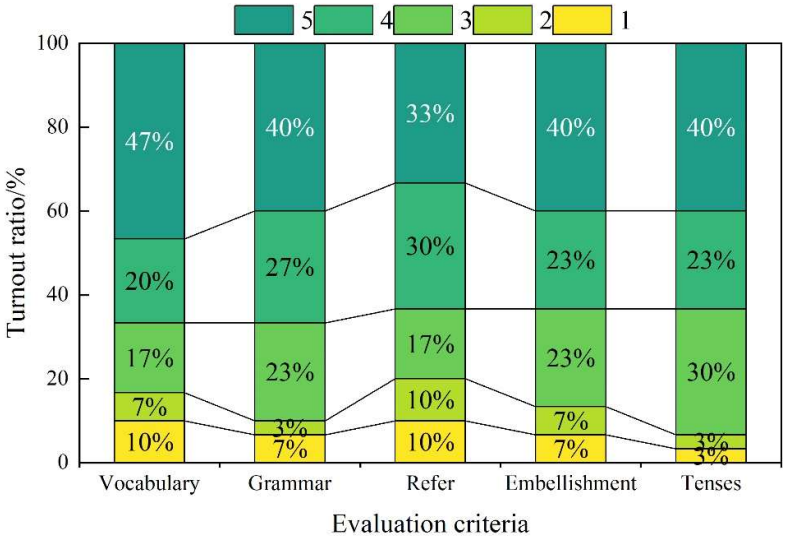

Fig. 8. The evaluation of the ambiguity of the translation

4.3. Case Studies

The study takes Wuthering Heights, a work of Victorian literature, as an example of translation using the method of this paper. The machine translated samples are also compared with Wu Guangjian's translation (T1), Liang Shiqiu's translation (T2), Rosset's translation (T3), Yang Coix's translation (T4), and Fang Ping's translation (T5).

4.3.1. Questionnaire design. Firstly, a comprehensive scoring method was adopted, in which the scorer scored each translated sentence according to the basic principles of style, emotional color and cultural connotation of the original text, taking into account the linguistic adaptability of the translation. Each translated sentence is followed by five scoring grades for selection, i.e., 2, 4, 6, 8 and 10 points. The higher the score, the better the linguistic adaptability of the translation, the better the quality of the translation. The level difference in the questionnaire is set at 2 points, which can facilitate the identification of gaps between translations and is easier for raters to operate.

4.3.2. Identification of raters. Although the comprehensive method can be used to examine the overall condition of the translations, it has some disadvantages. The raters are usually subjective, and the results of the survey will accordingly be mixed with many uncertainties, etc. In order to eliminate the influence of these factors, it is necessary to increase the number of samples, i.e., to use a larger number of raters. Moreover, the assessment is about the translation of literary works, so the raters need to have a high level of English-Chinese bilingualism and a certain understanding of translation standards. The raters selected for this study are English teachers (60%) or graduate students and doctoral students (40%) of English majors in colleges and universities, who have a high level of mastery of English-Chinese bilingualism.

A total of 110 paper questionnaires were distributed in this study and 100 were returned, with a recovery rate of 90.9%. Afterwards, data transcription of the survey results was performed and data integration and analysis were carried out using the statistical analysis tool SPSS 22.0.

4.3.3. Results and discussion. After the statistics, the average scores of the five translated sentences within each group in the 31 groups of sentences were all between 5.39 and 6.25, with the maximum mean value about 1 point higher than the minimum mean value, which is not a big difference. This

indicates that the raters were serious in the scoring process, the scores were more stable, and the credibility of the findings was high. Some statistical common measures were studied to further examine the concentration trend, degree of dispersion and distribution of the scores for each translation, etc.

A total of 5 score bands were set in the questionnaire, and the number of scores for each translation in each of the 5 bands is shown in Figure 9. From the figure, it can be seen that the number of scores of the six versions of translations in the five grades does not differ much, and the overall trend is more consistent. The number of translations translated based on the method of this paper with the highest score of 10 is 705, which is not much different compared to the manual translation.

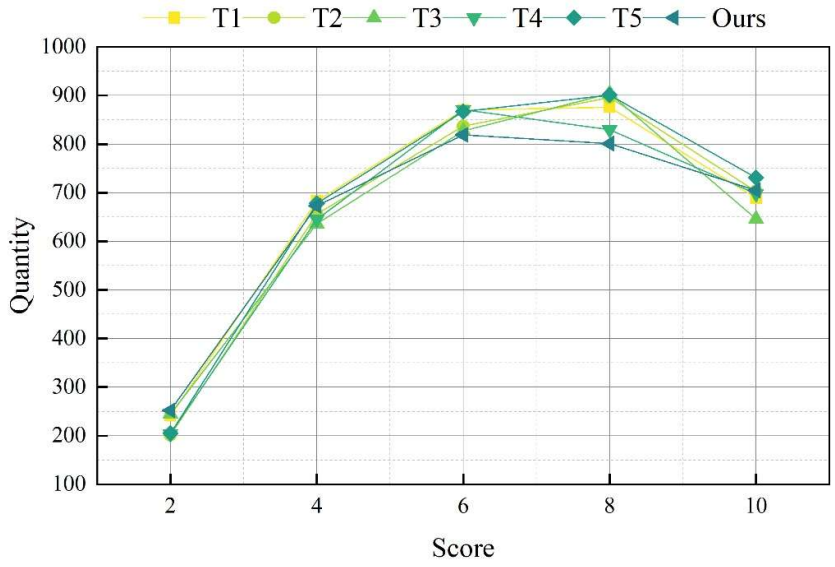

Fig. 9. The number of different points of score is counted in different translations

Through the statistical analysis of the questionnaire data to extract the eigenquantity indicators, including the mean, median, plural, standard deviation and other 8 items, which can illustrate the concentration trend, degree of dispersion and distribution of the translations' scores. The values of each statistical feature quantity are shown in Table 2. The maximum and minimum values reflect the distribution interval of the scores. The mean, median and plural can reflect the centralized trend and general level of the translation scores from different perspectives. The standard deviation reflects the discrete level (fluctuation size) of the translation scores. Skewness indicates the extent to which the distribution of translated scores deviates from the normal distribution (the larger the value, the more it deviates from the normal distribution). Kurtosis indicates the sharpness of the score distribution.

Whether it is human translation or machine translation, there are 2 and 10 points in the scores, indicating that there are some better and worse translations in both types of translations.The average value of the scores of the five human translations is 6.1, and the average score of the machine translations is 6.01, which indicates that the linguistic adaptability of translations based on the method of this paper is not much different from that of the human translations.

Table 2. Statistical characteristics of the translations

	T1	T2	T3	T4	T5	Ours
Min	2	2	2	2	2	2
Max	10	10	10	10	10	10
Mean	6.02	6.19	6.13	6.07	6.09	6.01
Median	6	6	6	6	6	6

Mode number	6	6	6	6	6	6
SD	2.267	2.27	2.238	2.228	2.227	2.247
Degree of bias	-0.128	-0.264	-0.191	-0.343	-0.185	-0.06
Kurtosis	-0.608	-0.652	-0.779	-0.772	-0.581	-0.731

5. Conclusion

In this paper, a character-level language modeling-based neural machine translation model incorporating text styles is designed for the translation of Victorian literature. The model under the text migration decoding module has the best performance among the three incorporation style representations, with the BLEU value, style migration accuracy, and style migration fluency reaching 37.49, 0.978, and 3.59, respectively.The model is used in the translation of Wuthering Heights literary works, and compared with Wu Guangjian's translation, Liang Shichu's translation, Rosset's translation, Yang Coixi's translation, and Fang Ping's translation. Evaluating the translations comprehensively from the three dimensions of original style, emotional color and cultural connotation, the average value of the scores of the five human translations is higher than that of the machine translation by 0.09. It shows that the machine translation model designed based on the language modeling technology in this paper has a better linguistic adaptability for translating Victorian era literary works.

Funding

2024 Guangxi Undergraduate Teaching Reform Project of Higher Education (Category A): Exploration and Practice on "Curriculum Ideology and Politics" of College English in Local Universities from the Perspective of High-Quality Development (Project No.: 2024JGA315).

References

[1] Saglia, D. (2018). European Literatures in Britain, 1815–1832: Romantic Translations: Romantic Translations (Vol. 123). Cambridge University Press.

[2] Tursunovich, R. I. (2022). Linguistic and cultural aspects of literary translation and translation skills. British Journal of Global Ecology and Sustainable Development, 10, 168-173.

[3] Usmonova, D. S., & Djalolov, Z. J. (2022, October). The role of culture in translation process. In International conferences (Vol. 1, No. 8, pp. 160-163).

[4] Morón, M., & Calvo, E. (2018). Introducing transcreation skills in translator training contexts: A situated project-based approach. The Journal of Specialised Translation, 29, 126-148.

[5] Kardiansyah, M. Y., & Salam, A. (2021, March). Reassuring feasibility of using Bourdieusian sociocultural paradigm for literary translation study. In Ninth International Conference on Language and Arts (ICLA 2020) (pp. 135-139). Atlantis Press.

[6] Kizi, A. D. R. (2023). Lexical errors and shortcomings in the translation process. European International Journal of Multidisciplinary Research and Management Studies, 3(10), 275-280.

[7] Jizzakh, A. Z. (2020). Marshak and Shakespeare: the question of new content in translation. Mental Enlightenment Scientific-Methodological Journal, 149-158.

[8] Volf, P. (2020). Translation techniques as a method for describing the results and classifying the types of translation solutions. Applied Translation, 14(2), 1-7.

[9] St. André, J. (2017). Metaphors of translation and representations of the translational act as solitary versus collaborative. Translation Studies, 10(3), 282-295.

[10] Toral, A., & Way, A. (2018). What level of quality can neural machine translation attain on literary text?. Translation quality assessment: From principles to practice, 263-287.

[11] Stahlberg, F. (2020). Neural machine translation: A review. Journal of Artificial Intelligence Research, 69, 343-418.

[12] Dabre, R., Chu, C., & Kunchukuttan, A. (2020). A survey of multilingual neural machine translation. ACM Computing Surveys (CSUR), 53(5), 1-38.

[13] Singh, S. P., Kumar, A., Darbari, H., Singh, L., Rastogi, A., & Jain, S. (2017, July). Machine translation using deep learning: An overview. In 2017 international conference on computer, communications and electronics (comptelix) (pp. 162-167). IEEE.

[14] Forcada, M. L. (2017). Making sense of neural machine translation. Translation spaces, 6(2), 291-309.

[15] Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and individual differences, 103, 102274.

[16] Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., ... & Xie, X. (2024). A survey on evaluation of large language models. ACM Transactions on Intelligent Systems and Technology, 15(3), 1-45.

[17] Wu, L., Zheng, Z., Qiu, Z., Wang, H., Gu, H., Shen, T., ... & Chen, E. (2024). A survey on large language models for recommendation. World Wide Web, 27(5), 60.

[18] Carlini, N., Tramer, F., Wallace, E., Jagielski, M., Herbert-Voss, A., Lee, K., ... & Raffel, C. (2021). Extracting training data from large language models. In 30th USENIX Security Symposium (USENIX Security 21) (pp. 2633-2650).

[19] Hou, X., Zhao, Y., Liu, Y., Yang, Z., Wang, K., Li, L., ... & Wang, H. (2023). Large language models for software engineering: A systematic literature review. ACM Transactions on Software Engineering and Methodology.

[20] Yuan, A., Coenen, A., Reif, E., & Ippolito, D. (2022, March). Wordcraft: story writing with large language models. In Proceedings of the 27th International Conference on Intelligent User Interfaces (pp. 841-852).

[21] Muskaan Singh,Ravinder Kumar & Inderveer Chana. (2020). Machine Translation Systems for Indian Languages: Review of Modelling Techniques, Challenges, Open Issues and Future Research Directions. Archives of Computational Methods in Engineering(4),1-29.

[22] Ma Kai,Zheng Shuai,Tian Miao,Qiu Qinjun,Tan Yongjian,Hu Xinxin... & Xie Zhong. (2023). CnGeoPLM: Contextual knowledge selection and embedding with pretrained language representation model for the geoscience domain. Earth Science Informatics(4),3629-3646.

[23] Mozhgan Ghassemiazghandi. (2024). An Evaluation of ChatGPT's Translation Accuracy Using BLEU Score. Theory and Practice in Language Studies(4),985-994.