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

Combinatorial Press

Research on Computational Methods for Optimizing the Teaching Content of Translation Based on Semantic Association Network Models

Peng Li[1], Yunxuan Zhang[2,✉]

[1] *School of Foreign Languages, Wuhan Polytechnic University, Wuhan, Hubei, 430048, China*

[2] *School of Foreign Languages, Wuhan City Polytechnic, Wuhan, Hubei, 430000, China*

ABSTRACT

Existing translation teaching content has certain deficiencies, this paper discusses the computational methods to optimize the translation teaching content by combining the semantic association network model. A domain translation model with joint semantic information is proposed, which constructs a bilingual mapping relation of domain-specific word vectors to obtain the semantic k-nearest neighbors of words in a specific domain,so as to estimate the domain intertranslation degree of words and improve the adaptive ability of the domain translation model. Then a semantic similarity computation model (SRoberta-SelfAtt) incorporating Robert's pre-training model is proposed. The model incorporates a self-attention mechanism to extract the association of different words within the text, and acquires richer sentence vector information. The proposed domain translation model is able to obtain more accurate translation results while spending less time. Compared with the stability of the iterative process of the basic model, the SRoberta-SelfAtt model has higher iterative stability. The Roberta-based semantic similarity computation model can effectively improve the performance of the word vector model. The experimental results show that the domain translation model with joint semantic information and the SRoberta-SelfAtt model are more practical for the task of optimizing translation teaching content.

Keywords: Semantic similarity computation, domain translation model, semantic k-nearest neighbor words, Robert pre-training model, translation teaching

✉ Corresponding author.
E-mail address: sevensisterbr@163.com (Y. Zhang).
Received 15 March 2024; Revised 01 May 2024; Accepted 15 November 2024; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-227

1. Introduction

Under the influence of globalization, the excellent cultures of various countries are undergoing the process of fusion by taking the essence and removing the meal, and English as the main language of international communication [1-2]. Attention is paid to the education of colleges and universities, and in the teaching of English in colleges and universities, teachers should pay attention to the practical use of English translation skills by students, so that the teaching of English translation has a purpose [3-4]. Teachers improve the quality of translation teaching is more helpful for students to further master a variety of social communication languages and improve the quality of conversation with foreign friends [5-6]. In actual communication, students can realize the feasibility of communication only when they can skillfully translate the meaning expressed by the other party, which also highlights the status of translation teaching [7-8].

Since the implementation of the new curriculum reform, the English teaching system is being optimized and improved, and universities have begun to pay attention to the improvement of students' knowledge application ability and intercultural communication ability [9-10]. Translation is one of the abilities that students must have, and it is also one of the basic abilities in the process of foreign language learning [11]. In specific teaching, teachers and colleges and universities should analyze the current problems of translation teaching and cultivate more high-quality and capable translators with the help of teaching situations, learning tasks, translation databases, curricula, practical activities and other ways [12-15].

Among them, in the process of cultivating translators, the linguistic translation strategy is the key to solve this problem. English-Chinese translation is a multi-attribute semantic decision-making problem, which needs to be analyzed by relevance and semantic similarity, combined with the fuzzy decision-making model to carry out the preferential design of semantic expression for translation [16-18]. The pragmatic translation strategy not only involves the transformation of linguistic forms, but also needs to take into account the communicative function of the translated message, including factors such as verbal style, implied meaning and context. Therefore, it is of great significance to explore the pragmatic translation strategies in English translation in colleges and universities [19-21].

Firstly, the semantic similarity problem in the translation process is outlined and analyzed. Literature [22] reveals the problems between languages in the language translation process, which mainly includes such problems as false equivalence of the translated languages, and argues that it is necessary to pay close attention to the correlation between the language translation process, and the consecutive translation of the utterances. Literature [23] devised a continuous measurement approach to investigate the problems associated with translation language misalignment and indirect mapping that reduces the accuracy of language translation, and based on the results of the study, it was learned that the relevance of translation choices significantly affects the one-to-one mapping built around the form and meaning of the translated text. Secondly, a method of translation word sense disambiguation and translation word sense similarity confirmation based on semantic association theme is explored to solve the above mentioned problems in the translation process. Literature [24] elaborated on the importance of correctly understanding the contextual semantics in the English translation process, so a method for determining the semantic similarity of English translation keywords based on the two-way LSTM neural network algorithm was conceived, and in the evaluation test, the similarity of the sentences translated based on the above method was approximate to the results of manual evaluation. Literature [25] developed a bilingual data-text generation and semantic analysis model, and the proposed system was evaluated and tested on WebNLG 2020 data, which confirmed the system's good performance in semantically related translation analysis. Literature [26] studied the problem of lexical disambiguation in the translation process, based on multimodal machine translation integrating visual information to enhance the purely translated text, and combining latent semantic analysis and sentence bert to extract the context vectors of English-speaking countries' corpora, and then realize

the evaluation of semantic diversity, and finally carried out a comprehensive evaluation in terms of the metrics such as BLEU, chrF2, and TER, which confirms that the proposed lexical disambiguation method effectively improves the quality of language translation.

The above methods of semantic disambiguation and semantic correlation are not only used to improve the quality of translation, but also can be introduced into the translation teaching classroom to improve the quality of translation teaching, and the current research on the optimization of translation teaching includes the feasibility of multimedia interactive computer-aided technology, corpus, neural network and other algorithmic technologies to empower translation teaching and the critical reflection on the empowerment of technology. Literature [27] analyzes the characteristics of English translation teaching in depth, proposes the introduction of data mining technology into English translation teaching to improve the effectiveness of English teaching translation, and carries out a detailed theoretical analysis in this regard. Literature [28] introduced computer-assisted translation (CAT) technology based on multimedia interaction into English translation teaching to improve students' English translation level, and built a machine translation model with fuzzy mathematics as the underlying logic to evaluate the teaching method, and finally explored the corresponding translation teaching program combined with the characteristics of CAT technology. Literature [29] introduces the construction, operation logic and scope of Multilingual Student Translation Corpus (MUST), which is suitable for teaching translation, designing translation materials and compiling teaching dictionaries. Literature [30] firstly learns and trains the real dataset and the public dataset PTB, constructs the SCN-LSTM translation model of deep learning neural network, and analyzes the performance of the translation model in translation teaching practice, and finds that the quality of translation teaching based on neural network is higher than that of the traditional n-tuple translation model. Literature [31] envisioned a CEFR-based mediation competence development online resource classification method for translation teaching, which enables the streamlining of resources for teaching second language communicative competence and translation competence, so that teachers of auxiliary language translation can be more flexible to the challenges of teaching English translation online. Literature [32] systematically reviewed the advantages and challenges of AI technology in the field of translation teaching, in which the lack of critical reflection on the hazards of AI by translation teaching stakeholders and the insufficient practice of related academic instructional concepts pointed to the need to further establish the principles of AI-based translation teaching and inspirational methods.

In this paper, a joint semantic information domain translation model optimization method is proposed, which uses the semantic distribution under the target domain to obtain the target mutual translation degree. Then a SRoberta-SelfAtt model incorporating Roberta's pre-training model is proposed to optimize the process of semantic similarity calculation for translating teaching content. The model is improved at three levels: text representation layer, feature encoding layer and similarity calculation layer. Subsequently, relevant experiments are designed to verify the model performance in conjunction with experimental evaluation metrics used for domain translation model validation, and finally the accuracy and iterative stability of the SRoberta-SelfAtt model proposed in this paper for calculating semantic similarity are compared with other models.

2. Domain translation model for joint semantic information

The domain translation model optimization method based on semantic similarity proposed in this paper aims at evaluating the domain intertranslatability of phrase pairs by using the semantic representations of words in the target domain in order to optimize the knowledge distribution of the generic translation model to adapt it to the target domain, and the structure of the method is shown in Fig. 1. The method is mainly divided into two main modules: 1) bilingual mapping of word vectors in the target domain. 2) computation of the domain intertranslatability of phrase pairs. Among them, the target domain word vector bilingual mapping module aims to obtain the cross-language semantic

k nearest neighbors of words in the target domain. The domain translation degree calculation module for phrase pairs is designed to estimate the domain translation degree of words based on the semantic distance between semantic k nearest neighbors and candidate words in the universal domain, and then obtain the domain translation probability of phrase pairs.

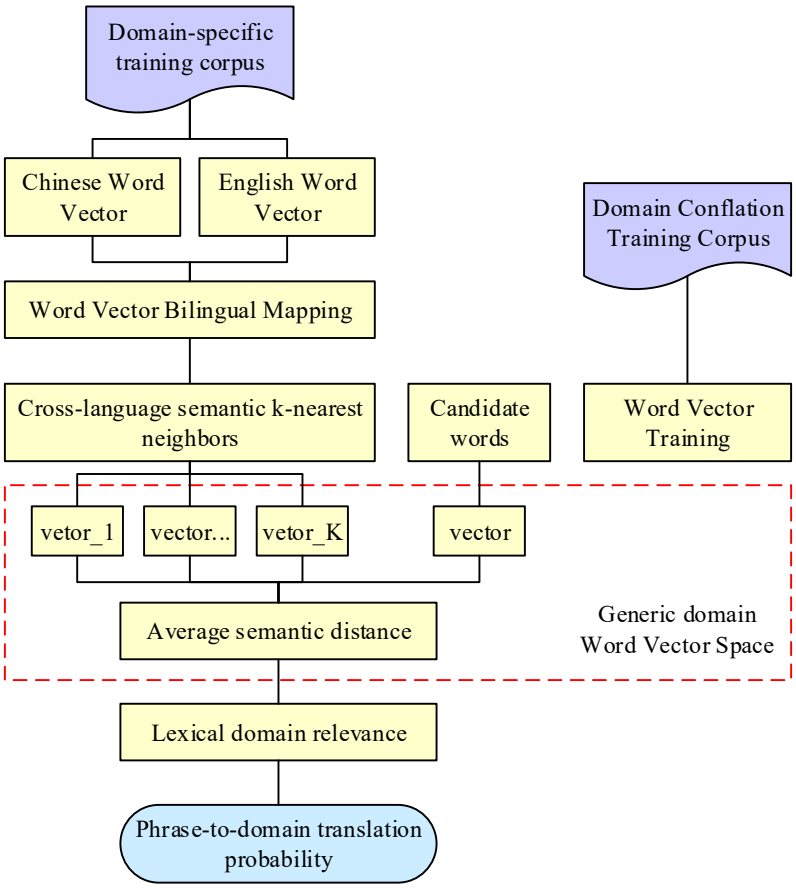

Fig. 1. Based on semantic similarity optimization method

2.1. Neural network-based bilingual mapping of domain word vectors

This paper proposes to use neural network modeling to construct the bilingual mapping relations of word vectors in the target domain, which aims at obtaining the cross-language semantic representation of words in the target domain. To construct the training corpus, this paper first selects the set of words with the highest frequency of occurrence from the target domain, and with the help of the word alignment relation of the parallel corpus, finds out the translation with the highest probability of intertranslation with the word in order to construct the target domain dictionary. Second, the mutually translated word pairs in the bilingual dictionary are characterized into corresponding domain word vector pairs, which are used as the training data for the neural network model.

The input layer of the neural network model is the word vectors of the source language words, which arrive at the output layer after two nonlinear transformations to get the mapping results of the word vectors in the target language end. The training goal of the neural network model is to minimize the sum of the errors between the mapping results of the output layer and the target word vectors. As equation (1) gives the formula for the neural network input layer to the output layer, and equation (2) gives the mathematical definition of the loss function of the neural network model.

$$o = g\left(W_2\left(g\left(W_1 x + b_1\right) + b_2\right)\right) \tag{1}$$

$$E=\frac{1}{2}\sum_{d=5}\left(t_d-o_d\right)^2 \tag{2}$$

Where, $g(x)$ denotes the activation function, in this paper we use the sigmoid function as the activation function, W_1 denotes the parameter matrix from the input layer of the neural network to the hidden layer. W_2 denotes the parameter matrix from the hidden layer to the output layer of the network. b_1, b_2 denote the bias terms. S denotes the training data set, t_d denotes the word vector of the target word in the d th training sample, and o_d denotes the mapping result of the word vector of the source language in the d th sample. In this paper, the stochastic gradient descent algorithm is used to learn the network parameters.

2.2. Evaluation Model of Interpretation Degree of Phrase-to-Domain

In this paper, we utilize a parallel corpus with large scale and mixed domains to extract the translation phrase table, and calculate the mutual translation probabilities of all phrase pairs in the phrase table under the target domain, which are finally added as new domain features to the translation system. In the initial stage, this paper utilizes target domain and generic domain monolingual resources to pre-train monolingual word vectors of different languages in the corresponding domain. Second, for a phrase pair, this paper obtains the set of all inter-translated word pairs based on the word alignment relations within the phrase pair. Then, for each word pair, their domain intertranslation degree is calculated based on the semantic distribution distance of the corresponding word vectors; finally, based on the domain intertranslation degree of the word pairs, the intertranslation probability of the phrase pairs in the target domain is estimated.

1) Calculation of word pairs' mutual translation degree. In order to calculate the domain intertranslation degree of word pairs, this paper utilizes the bilingual mapping method of domain word vectors to map the source language words to the target language side, and adopts the cosine similarity computation to obtain the k word that is semantically most relevant to the mapping vector, called the semantic k nearest neighbor word.

First, this paper calculates the semantic relevance of semantic k nearest neighbor words and target candidate translations. Since the semantic k nearest neighbor words and the target candidate translations belong to the same language, this paper directly calculates their cosine distances in the generic domain word vector space as their semantic relatedness.

Secondly, for the source language word, this paper estimates its intertranslation degree with all its semantic k nearest neighbors, which is obtained by calculating the cosine distance between the mapping vectors of the word and the word vectors of a certain k nearest neighbor, and normalized using the sum of the intertranslation degrees with all the semantic k nearest neighbors in order to satisfy the property that the sum of the probability of the word pairs' intertranslations is one. The specific calculation method is shown in Equation (3):

$$t\left(e^{(k)}\mid f_i\right)=\frac{\exp\left\{sim\left(e^{(k)}\mid e'\right)\right\}}{\sum_{e^{(k)}\in S(e^*)}\exp\left\{sim\left(e^{(k)}\mid e'\right)\right\}} \tag{3}$$

where $t\left(e^{(k)}\mid f_i\right)$ denotes the degree of intertranslation between word f_i and its semantic k immediate neighbors $e^{(k)}$ at the target end, e' denotes the mapping vector of word f_i at the target language end, and $S\left(e'\right)$ is the set of semantic k immediate neighbors of word f_i at the target language end.

Finally, the joint model is utilized to fuse the two scores, which is used to estimate the translation probability of source language words and target candidate words in the target domain. The specific calculation method is shown in Equation (4):

$$p\left(e_j \mid f_i\right)=\sum_{e_k \in\left(e^*\right)} sim\left(e_j, e^{(k)}\right) t\left(e^{(k)} \mid f_i\right) \tag{4}$$

where $p\left(e_j \mid f_i\right)$ denotes the probability of translation between the source language word and the target word in the target domain. $sim\left(e_j, e^{(k)}\right)$ denotes the semantic correlation between the semantic k nearest neighbor word e_k and the target word e_j.

2) Phrase pair intertranslation degree calculation. In this paper, we estimate the intertranslation probability between phrase pairs under a specific domain based on the lexicalized weighting formula of phrases. First, this paper calculates the probability that a source language word is translated into a target language word. Secondly, based on the lexicalization weighting formula, the positive phrase inter-translation probability is calculated and added as a new domain feature to the translation system. The calculation method is shown in Equation (5) below:

$$Similarity\left(\bar{e} \mid \bar{f}, a\right)=\prod_{j=1}^{langth(\bar{e})} \frac{1}{\left|\{i \mid (i, j) \in a\}\right|} \sum_{(i, j),, a} p\left(e_j \mid f_i\right) \tag{5}$$

In order to better evaluate the domain intertranslation degree of phrase pairs, this paper calculates the intertranslation probability of phrase pairs from the opposite direction. Similarly, this paper constructs the mapping relation from target language to source language word vector, and obtains the semantic k nearest neighbor word of the target word in the source language. Based on this semantic k nearest neighbor word, this paper calculates the probability of translating a word at the target end into a word at the source end. Then, using the lexicalization weighting formula, the reverse phrase domain intertranslation degree is calculated and added as a new domain feature into the translation system. The specific calculation method is shown in Equation (6):

$$Similarity\left(\bar{f} \mid \bar{e}, a\right)=\prod_{i=1}^{length(\bar{f})} \frac{1}{\left|\{j \mid (i, j) \in a\}\right|} \sum_{(i, j) \in a} p\left(f_i \mid e_j\right) \tag{6}$$

where $\bar{f}$ denotes the source language phrase, $\bar{e}$ denotes the target language phrase, a denotes the word alignment relation within the phrase pair, and $length(f)$ and $length(e)$ denote the lengths of the source language phrase and target language phrase e, respectively. $p\left(e_j \mid f_i\right)$ and $p\left(f_i \mid e_j\right)$ denote the domain intertranslation probabilities of f_i and e_j in the forward and backward directions, respectively.

3. Semantic similarity calculation based on Roberta's translation teaching content

The word vectors trained by the large-scale pre-training model can dynamically characterize the word vectors according to the different contexts of the text, so that the twin network structure can extract the richer semantic features of the text at the input layer and effectively reduce the complexity of the next model layer. A representative model architecture is the SBERT structure [33].

The model architecture is intuitive and simple, and although it has been experimentally verified to have achieved good results, it has not fully extracted the correlation relationship between words and the interaction information between texts. In order to solve this problem, this chapter proposes an improved semantic similarity computation model SRoberta-SelfAtt for English short texts based on the above model.

Figure 2 shows the structure of SRoberta-SelfAtt model, which contains three layers: text representation layer, feature encoding layer and similarity computation layer, and the improvement points of this chapter focus on the text encoding layer and feature representation layer. For the text representation layer, the pre-trained model of Roberta [34] is chosen instead of BERT [35] to better represent the original text as word-level vectors. For the feature encoding layer, this chapter obtains

the correlation information between different words in the text through the self-attention mechanism, and then obtains the sentence vector information from different perspectives through the global maximum pooling strategy and the global average pooling strategy, and then fuses the two different sentence vectors to ensure that the final sentence vectors can fully represent the original text.

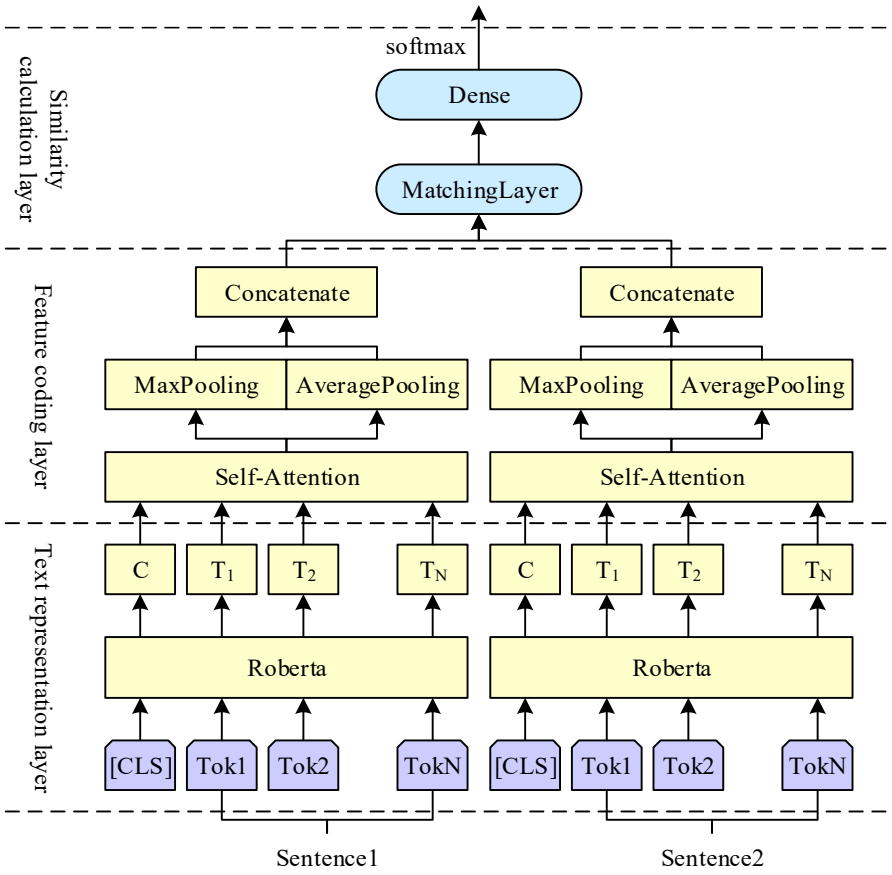

Fig. 2. Sroberta-selfatt model

3.1. Text representation layer

The Roberta pre-training model optimizes the pre-training task of BERT without changing the original structure of BERT, so that the text representation layer of the SRoberta-SelfAtt model consists of two identical Roberta pre-training models.The main improvements of the Roberta model compared to BERT include the following four aspects:

1) Longer training model time, larger training batches and more training data. The original BERT shortens the training time of the model by randomly injecting short sequences and shortening the sequence length for the first 90% of the sequences in the model training, while Roberta only trains full-length sequences.

2) Removing the next sentence prediction task. For the next sentence prediction task, positive samples are taken from consecutive sentences in the corpus, while negative samples are created by pairing sentence segments from different documents, and both positive and negative samples are sampled with the same probability. The goal of this task is to predict whether two text segments in the original text are connected or not, and it is commonly used for performance improvement in downstream tasks such as natural language reasoning.

3) Replacing static masks with dynamic masks in pre-training tasks. In the BERT source code, random masking and substitution is performed only once at the beginning and saved during training, thus, it can be regarded as static masking. Pre-training of BERT relies on random masking and prediction of

masked words or phrases. To avoid using the same mask for each training instance in each epoch, the training data is repeated 10 times.

4) Use of character encoding (BPE). Character encoding is a hybrid between character-level and word-level representations, which can handle large vocabularies commonly found in natural language corpora and avoid more "[UNK]" sign symbols in the training data, which can affect the performance of the pretrained model.

Roberta model is based on the BERT structure, Input is the input of the original text, and the special symbol "[CLS]" at its starting position indicates the vector of the whole sentence. The Embedding of the model consists of three parts, namely Token Embedding, Segment Embedding and Position Embedding. Assuming that the input text including the special symbol "[CLS]" is the maximum length of the input sentence set by the model, the inputs to the Roberta model will each be encoded as vectors:

$$E_i = E_{token}\left(x_i\right) + E_{seg}\left(x_i\right) + E_{pos}\left(x_i\right) \tag{7}$$

Where: $E_{token}\left(x_i\right)$, $E_{seg}\left(x_i\right)$ and $E_{pos}\left(x_i\right)$ denote the Token Embedding, Segment Embedding and Position Embedding vector information respectively, after which $\left[E_1, E_2, \ldots, E_n\right]$ is used as the input to the bidirectional Transformer network. Transformer is essentially an encoder-decoder model, and it is the coding unit module in Transformer that Roberta utilizes to encode the input Embedding information, load the pre-training completed parameters, and arrive at the final vector representation of the input sequence $\left[T_1, T_2, \ldots, T_n\right]$.

3.2. Feature coding layer

The feature encoding layer firstly captures the correlation and importance of different words in the text through the self-attention mechanism, and selects the more critical information for the current task goal from the many words in the text, and then extracts the sentence vector information through the two strategies of global maximum pooling and global average pooling, and splices them together to obtain the final vector representation of the text. The principle of the self-attention mechanism is as follows:

When the human eye observes a picture, it pays different attention to different areas of the picture, and its visual attention will instinctively pay attention to the areas that have the most critical impact on the information in the picture, so that the most valuable information can be quickly filtered out, which is where the idea of the attention mechanism comes from. The self-attention mechanism aims to reduce the influence of external information, and is more adept at capturing the internal relevance of data and features. In natural language processing, after the pre-training model vectorized representation of the text will generate a word vector matrix, to do Self-Attention that is, for each word in the text, are queried the degree of match between the current word and each contextual word, the degree of its size will affect the number of features added. The Attention mechanism is calculated as shown in equation (8):

$$Attention\left(Q, K, V\right) = soft\max\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{8}$$

The calculation steps are:

Step1: Multiply the text word vector matrix by the three weight matrices W^Q, W^K and W^V respectively to get *query*, *key* and *value*, denoted as Q, K, V, where Q, K, V has the same dimensions as $N \times d_k$, $N \times d_k$, $N \times d_v$, Q and K respectively.

Step2: Calculate the inner product of Q and every K of the current word of the text, i.e. $\frac{QK^T}{\sqrt{d_k}}$, where d_k is a penalty factor to avoid QK^T being too large. If linearly irrelevant, the inner product is 0. The larger the relevance, the larger the inner product, the closer the degree of match between the

two words. The size of the matching degree determines the size of the influence of the context word on the current word.

Step3: Combine the calculated inner product of Step2 with the calculation results of each V to form the final feature expression of the current word $\left(\frac{QK^T}{\sqrt{d_k}}\right)V$.

Step4: Combine the calculation results of Step3 with the $soft\max$ function to get the final vector representation of the current word incorporating contextual information. $soft\max$ is the normalized exponential function, which compresses the elements of the row vector to between $[0,1]$ and the vector elements sum to 1.

3.3. Similarity calculation layer

The similarity calculation layer obtains the interaction information between text pairs through a customized matching network Matching Layer, assuming that the sentence vectors encoded by the feature coding layer are u and v respectively, the calculation formula of Matching Layer is:

$$MatchingLayer = Concatenate\left(u, v, |u - v|\right) \tag{9}$$

Then the vectors obtained from MatchingLayer are input into the fully connected layer, using softmax as the activation function, and adopting the categorical cross entropy categorical_crossentropy as the loss function, which is calculated by the formula:

$$Loss = -\sum_{i=1}^{n} y_{(i,m)} \log\left(\hat{y}_{i,m}\right) \tag{10}$$

Where: $y_{i,m}$ denotes the probability that the sample in category i of the true value belongs to category m, while $\hat{y}_{i,m}$ denotes the probability that the sample in category i of the predicted value belongs to category m.

4. nglish Translation Effect Analysis

4.1. Experiment preparation phase

In order to verify the English translation effect of the proposed method, several groups of English teaching passages are selected as experimental objects. Since the translation model training parameters are more important in the proposed method, the optimal parameters for translation model training are determined before the experiment is conducted as a way to simplify the experimental steps and time. It is known that the auxiliary parameters for model relevance calculation are generally between [1, 10], and χ and δ represent the translation model training parameters. The relationship between the value of χ and the accuracy of the translation model is calculated when δ is set to 5, and the results obtained are shown in Figure 3. When δ is 5, the translation model accuracy reaches the maximum value of 88.9% when χ takes the value of 5.

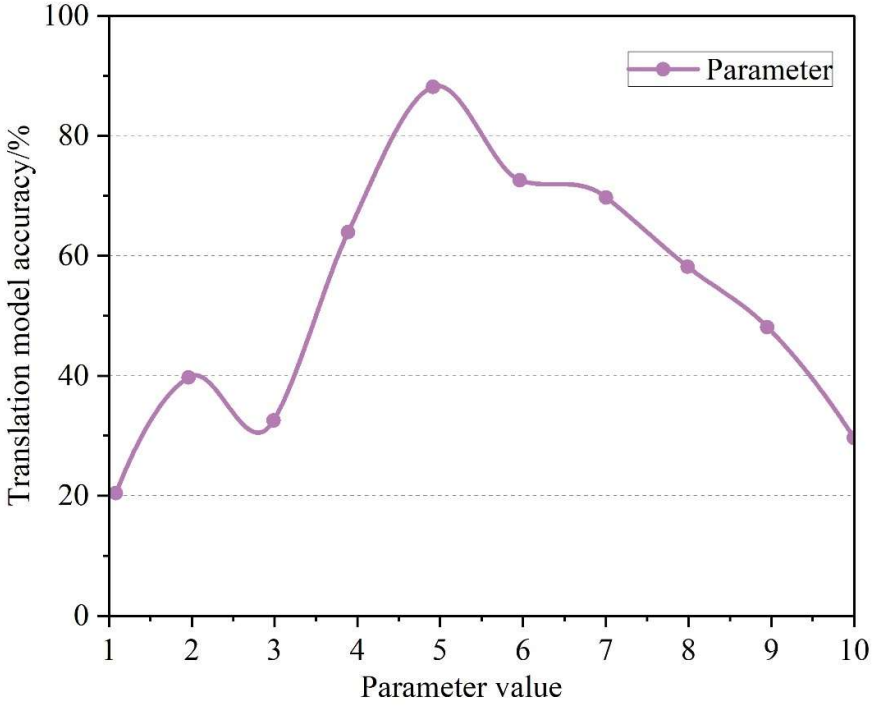

Fig. 3. Parameters and translation model precision relationships

Taking χ as 5, the relationship between δ and translation model accuracy is calculated and the results are obtained as shown in Figure 4. When the parameter χ is 5, δ takes the value of 3 translation model accuracy reaches the maximum value of 77.8%. Then the optimal parameters for translation model training are determined as χ=5, δ=3. The above process completes the determination of the optimal parameters for translation model training, which facilitates the smooth progress of the subsequent experiments.

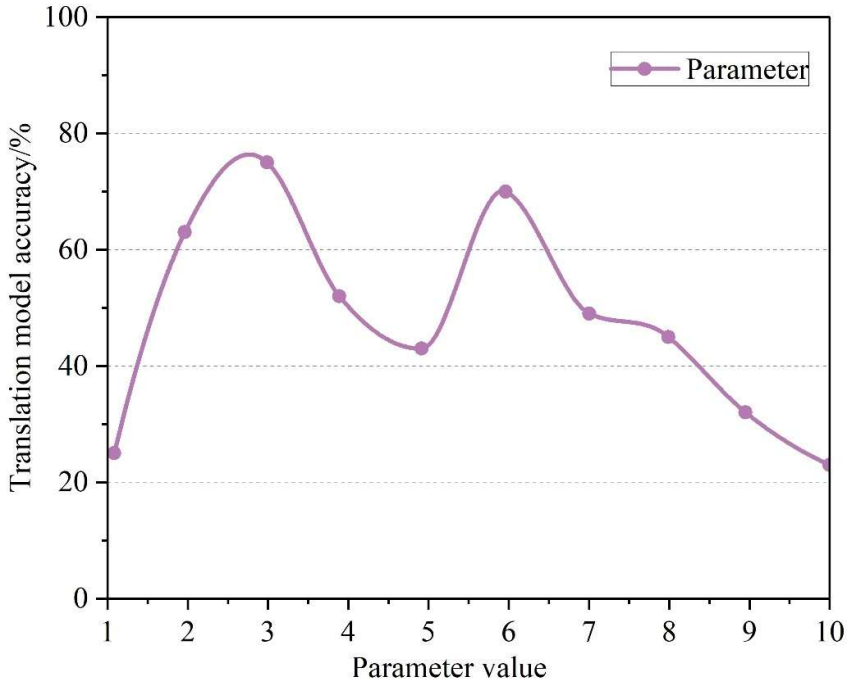

Fig. 4. Parameters and translation model precision relationships

4.2. Selection of evaluation indicators

In order to visualize the application performance of the proposed method, English translation time and BLEU index are selected as evaluation indexes, and the calculation formula is:

$$\begin{cases} T = t_2 - t_1 \\ BLEU = BP \times \exp\left(\sum_{n=1}^{N} a_n \log I_n\right) \end{cases} \tag{11}$$

In Eq. (11), T represents the English translation time; t_1 and t_2 represent the English paragraph translation request time and feedback time; BP represents the length penalty factor; and N represents the predetermined longest tuple. a_n is the weight corresponding to the English translation accuracy. I_n is the probability that the n th word is translated correctly.

Conventionally, the shorter the English translation time and the closer the BLEU index (with a value range of $[0,1]$) is to 1, the better the English translation is.

4.3. Analysis of experimental results

A certain English teaching article was selected as the data set for the experiment, and the English translation experiment was conducted on it based on the optimal training parameters determined above and the evaluation indexes selected. Based on the experimental results, the evaluation indexes (English translation time and BLEU index) were statistically calculated. The English translation time data obtained through the experiment is shown in Figure 5. Under different experimental conditions, the English translation time obtained by applying the proposed method is lower than the maximum limit, and the minimum value reaches 0.38s, indicating that the proposed method is more efficient in English translation.

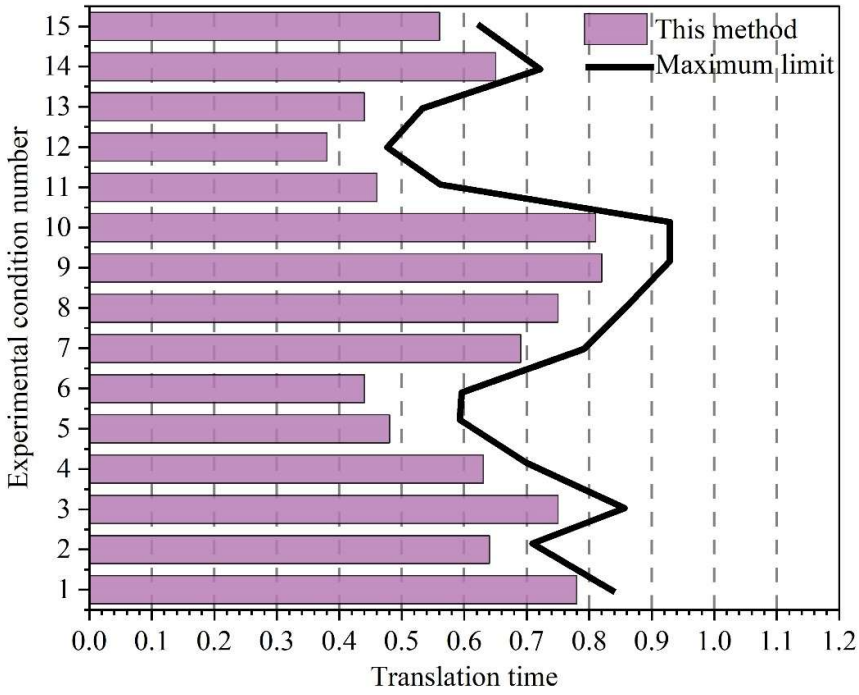

Fig. 5. English translation time data diagram

The BLEU indices obtained through the experiments are shown in Table 1. The BLEU indices obtained by applying the proposed method are all higher than the standard values, and the maximum value reaches 0.93, indicating that the proposed method is more effective in English translation. The above experimental results show that compared with the given maximum value and standard value,

the English translation time obtained by applying the method of this paper is shorter and the BLEU index is larger, which fully confirms that the proposed method of this paper has a better English translation effect.

Table 1. BLEU indicator table

Experimental condition number	Proposed method	Standard value
1	0.8	0.55
2	0.9	0.75
3	0.89	0.69
4	0.87	0.72
5	0.92	0.67
6	0.86	0.8
7	0.76	0.7
8	0.79	0.75
9	0.92	0.65
10	0.85	0.7
11	0.73	0.72
12	0.93	0.8
13	0.74	0.68
14	0.76	0.7
15	0.92	0.64

5. Effectiveness Analysis of Semantic Similarity Calculation of Translation Teaching Content

In order to demonstrate the effectiveness of the proposed improved SRobert-SelfAtt model, we conducted three groups of comparison experiments, and the results are shown in Table 2, No.1 represents the experimental results of the standard SBERT model before optimization, No.2 represents the experimental results of the SBERT model which only incorporates two semantic constraints: tautology and antithetical, and No.3 represents the experimental results of the improved SRobert- SelfAtt model, which incorporates information from two different sentence vectors.The results of group No.2 are 0.526ρ/0.495 r, which are 0.171ρ and 0.136r higher than those of the standard SBERT model before optimization, and the results of the improved SRobert-SelfAtt model in group No.3 reach 0.618ρ/0.522 r, which are higher than those of the optimized standard SBERT model, while the results of group No.3 reach 0.618ρ/0.522 r, which are higher than the optimized standard SBERT model. 0.522 r, which are 0.263ρ and 0.027r higher than the results of the standard SBERT model before optimization, respectively.It is easy to see that the improved SRobert-SelfAtt model is very effective in enhancing the existing word vector model, and it can effectively improve the basic standard SBERT model.

Table 2. Comparison of experimental results of three methods

No.	Method	Spearman ρ	Pearson r
1	SBERT	0.355	0.359
2	Integrated into two semantic SBERT models	0.526	0.495
3	SRobert-SelfAtt	0.618	0.522

Figure 6 shows the learning curves of the basic standard SBERT model and the improved SRobert-SelfAtt model within 20 iterations. It is easy to see that both models reach the convergence state within

20 iterations, which indicates that the two methods can find feasible solutions within a finite number of iterations, and the results of the improved model are significantly better than those of the basic model after reaching the convergence state. Secondly, after the 7th iteration, the performance of the two methods is close to their own optimal performance, which indicates that the short period iteration can reach the vicinity of their corresponding maximum values, and therefore, the two models also have some advantages in terms of computing time. In addition, the performance of the basic model shows a slight degradation during 10-20 iterations, compared to the improved model, which shows a more stable performance.

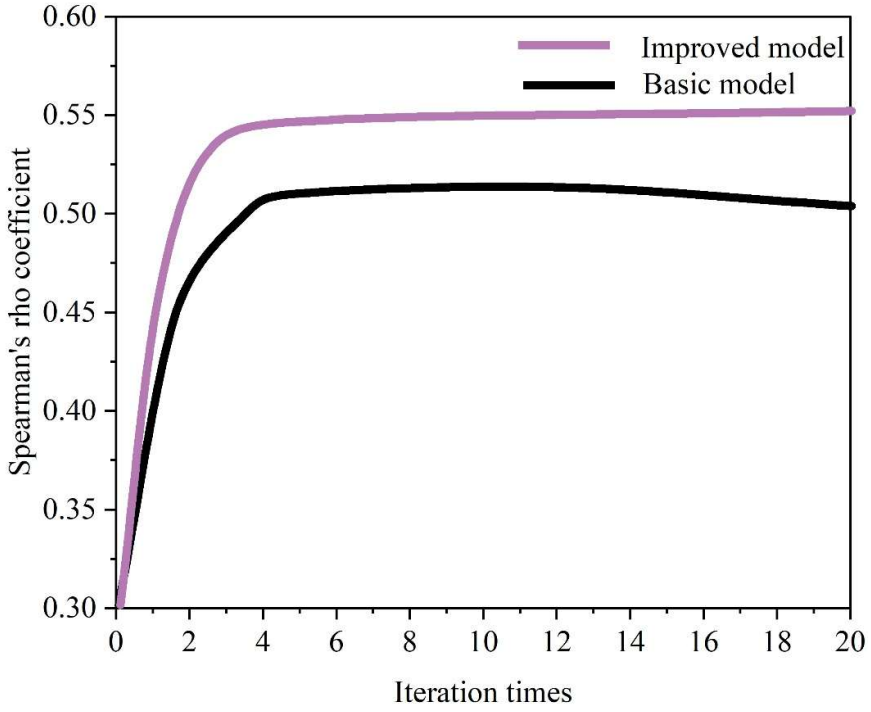

Fig. 6. The learning curve of two models

In order to further analyze the experimental results, 10 pairs of examples were selected, and Table 3 shows the contents of the selected examples, which can clearly show the improvement of the SRobert-SelfAtt method, and its corresponding word vectors were visualized and displayed in Figures after dimensionality reduction by the principal component analysis algorithm, and the distributions of the word pairs before and after the improvement are shown in Figures 7 and 8, respectively. The gap between the overall word pair scores and the manual scores before the improvement is large, while the gap between the scores of the word pairs after the improvement is very small, and the gap between the scores of the word pair numbered 10 after the improvement and the manual scores is about 1%.

Table 3. The Top 10 Word Pairs Change Better after Counter-fitting

No.	Word pair	Manual scoring	Improvement	improved
1	listen/hear	9.6	6.619	9.886
2	class/lesson	9.2	6.127	7.619
3	everyone/everybody	8.5	5.315	9.843
4	glass/cup	8.7	5.627	10.000
5	large/big	9.4	6.357	9.997
6	glad/happy	9.5	6.455	10.000

7	like/love	3.3	7.587	6.58
8	little/small	3.1	7.692	6.924
9	start/begin	9.4	6.899	9.999
10	near/beside	9.9	7.402	9.998

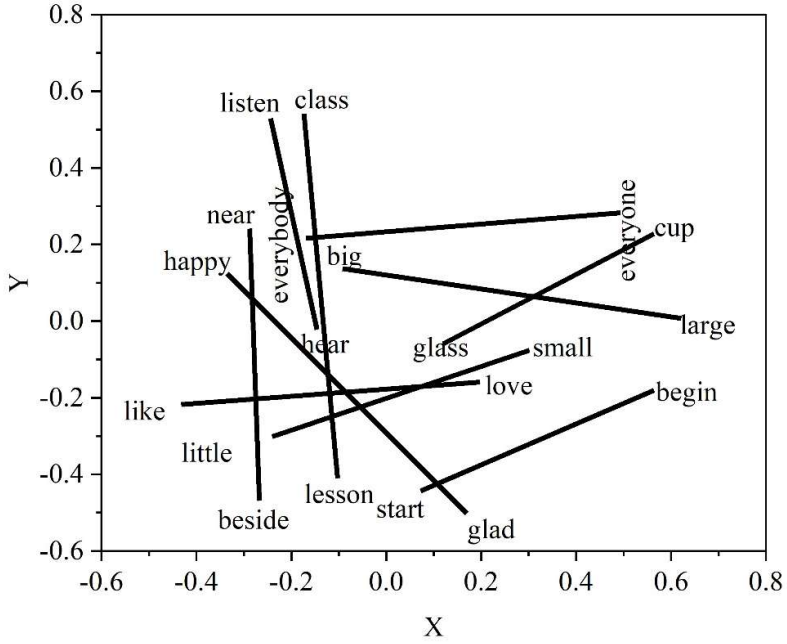

Fig. 7. The word is improved before the distribution

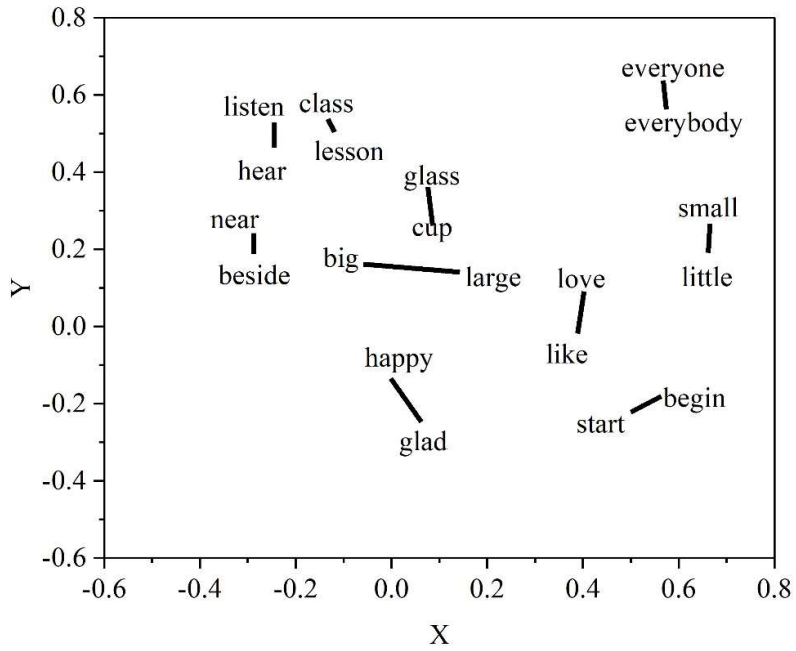

Fig. 8. The improved words are distributed

6. Conclusion

Aiming at the current translation teaching still exists problems such as imperfect content of teaching materials, insufficient teachers and lagging teaching methods, this paper analyzes and summarizes the characteristics of the existing methods, and proposes a similarity calculation method for translation

teaching content jointly with semantic relational network model and deep learning model.

The proposed domain translation model can obtain accurate English translation results and spend less time, and the time minimum is 0.38s, which shows the high efficiency of English translation under the model. The corresponding BLEU indices are all larger than the standard values, and the maximum value reaches 0.93. It confirms that the domain translation model with joint semantic information proposed in this paper has better English translation results and provides an effective method support for English translation.

The Roberta-based semantic similarity computation model improves 0.263ρ and 0.027r over the standard SBERT model before optimization. The basic model shows a slight performance degradation during the iteration process, while the model in this paper iterates smoothly. It can be seen that the Roberta-based semantic similarity computation model can effectively improve the performance of the word vector model in translation teaching content.

References

[1] Li, L. (2021). English translation teaching model of flipped classroom based on the fusion algorithm of network communication and artificial intelligence. Wireless Communications and Mobile Computing, 2021(1), 7520862.

[2] Ismatullayeva, N. R. (2022). Teaching Translation Methodology in the Foreign Language Classes. International Journal of Social Science Research and Review, 5(12), 448-452.

[3] Krüger, R. (2021). AN ONLINE REPOSITORY OF PYTHON RESOURCES FOR TEACHING MACHINE TRANSLATION TO TRANSLATION STUDENTS. Current Trends in Translation Teaching & Learning E.

[4] Mažeikienė, V. (2019). Translation as a method in teaching ESP: An inductive thematic analysis of literature. Journal of Teaching English for Specific and Academic Purposes, 513-523.

[5] Omar, A., & Gomaa, Y. (2020). The machine translation of literature: Implications for translation pedagogy. International Journal of Emerging Technologies in Learning (iJET), 15(11), 228-235.

[6] Cherchata, L., Korol, L., Rubchak, O., Orda, O., & Novytska, D. (2023). Effectiveness of translation transformations in different styles of the english language for teaching written translation. Amazonia Investiga, 12(67), 185-197.

[7] Benites, A. D., Kureth, S. C., Lehr, C., & Steele, E. (2021). Machine translation literacy: a panorama of practices at Swiss universities and implications for language teaching. CALL and professionalisation: short papers from EUROCALL 2021, 80.

[8] Kübler*, N., Mestivier, A., & Pecman*, M. (2018). Teaching specialised translation through corpus linguistics: translation quality assessment and methodology evaluation and enhancement by experimental approach. Meta, 63(3), 807-825.

[9] Yan, D., & Wang, J. (2022). Teaching data science to undergraduate translation trainees: Pilot evaluation of a task-based course. Frontiers in Psychology, 13, 939689.

[10] Ma, F. (2022). Application of convolutional neural network based on deep learning in college english translation teaching management. Mobile Information Systems, 2022(1), 9673133.

[11] Pan, B., & Qin, Q. (2022). Construction of parallel corpus for english translation teaching based on computer aided translation software. Computer-Aided Design and Applications, 19(1), 70-80.

[12] Gonzales, L. (2018). Sites of translation: What multilinguals can teach us about digital writing and rhetoric (p. 153). University of Michigan Press.

[13] Wei, X. (2019). Research on the Development Strategy of Cross-cultural Ecological Education in English Translation Teaching at Colleges. Ekoloji Dergisi, (107).

[14] Zhu, L. (2018). An embodied cognition perspective on translation education: philosophy and pedagogy. Perspectives, 26(1), 135-151.

[15] Benelhadj Djelloul, D., & Neddar, B. A. (2017). The usefulness of translation in Foreign language teaching: Teachers' attitudes and perceptions. AWEJ for translation & Literacy Studies Volume, 1.

[16] Hoang, J. (2019). Translation technique term and semantics. Applied Translation, 13(1), 16-25.

[17] Mohamed, E., Elmougy, S., Ibrahim, O. A. S., & Aref, M. (2019). Semantic relatedness based query translation disambiguation approach for cross-language web search. Int J Adv Sci Technol.

[18] Xiao, Y., Keung, J., Bennin, K. E., & Mi, Q. (2018). Machine translation-based bug localization technique for bridging lexical gap. Information and Software Technology, 99, 58-61.

[19] Praet, S., & Verhelst, B. (2020). Teaching translation theory and practice. Journal of Classics Teaching, 21(42), 31-35.

[20] Liang, Y. (2021, February). English character image feature semantic block processing for English-Chinese machine translation. In 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV) (pp. 841-845). IEEE.

[21] Siregar, R. (2018). Grammar based translation method in translation teaching. International Journal of English Language and Translation Studies, 6(02), 148-154.

[22] Alisherovna, X. M. (2024). SEMANTIC ASPECTS OF TRANSLATION. International journal of advanced research in education, technology and management, 3(9), 6-12.

[23] Bracken, J., Degani, T., Eddington, C., & Tokowicz, N. (2017). Translation semantic variability: How semantic relatedness affects learning of translation-ambiguous words. Bilingualism: Language and Cognition, 20(4), 783-794.

[24] Zhili, W., & Qian, Z. (2024). A Deep Learning-based Method for Determining Semantic Similarity of English Translation Keywords. International Journal of Advanced Computer Science & Applications, 15(5).

[25] Agarwal, O., Kale, M., Ge, H., Shakeri, S., & Al-Rfou, R. (2020). Machine translation aided bilingual data-to-text generation and semantic parsing. In Proceedings of the 3rd international workshop on natural language generation from the semantic web (WebNLG+) (pp. 125-130).

[26] Hatami, A., Arcan, M., & Buitelaar, P. (2024, September). Enhancing Translation Quality by Leveraging Semantic Diversity in Multimodal Machine Translation. In Proceedings of the 16th Conference of the Association for Machine Translation in the Americas (Volume 1: Research Track) (pp. 154-166).

[27] Chen, J. (2020, December). Strategies for improving the effectiveness of English translation teaching in higher vocational colleges based on data mining. In Journal of Physics: Conference Series (Vol. 1693, No. 1, p. 012021). IOP Publishing.

[28] Wu, H. (2021). Multimedia Interaction - Based Computer - Aided Translation Technology in Applied English Teaching. Mobile Information Systems, 2021(1), 5578476.

[29] Granger, S., & Lefer, M. A. (2020). The Multilingual Student Translation corpus: a resource for translation teaching and research. Language Resources and Evaluation, 54(4), 1183-1199.

[30] Ren, B. (2020). The use of machine translation algorithm based on residual and LSTM neural network in translation teaching. Plos one, 15(11), e0240663.

[31] Enbaeva, L., & Plastinina, N. (2021). Distance learning challenges in translation teaching: Mediation competence development. In E3S Web of Conferences (Vol. 258, p. 07078). EDP Sciences.

[32] Liu, K., & Afzaal, M. (2021). Artificial Intelligence (AI) and translation teaching: A critical perspective on the transformation of education. International journal of educational sciences, 33(1-3), 64-73.

[33] Burcu Yalçıner,Kıvanç Dinçer,Adil Gürsel Karaçor & Mehmet Önder Efe. (2024). Enhancing Agile Story Point Estimation: Integrating Deep Learning, Machine Learning, and Natural Language Processing with SBERT and Gradient Boosted Trees. Applied Sciences(16),7305-7305.

[34] Yuezhong Wu,Huan Xie,Lin Gu,Rongrong Chen,Shanshan Chen,Fanglan Wang... & Jinsong Tang. (2024). Advancing Mental Health Care: Intelligent Assessments and Automated Generation of Personalized Advice via M.I.N.I and RoBERTa. Applied Sciences(20),9447-9447.

[35] Mohan K. Mali,Ranjeet R. Pawar,Sandeep A. Shinde,Satish D. Kale,Sameer V. Mulik,Asmita A. Jagtap... & Punam U. Rajput. (2025). Automatic detection of cyberbullying behaviour on social media using Stacked Bi-Gru attention with BERT model. Expert Systems With Applications125641-125641.