ENGLISH MACHINE TRANSLATION BASED ON WORD SLICING ALGORITHM

Shuhan YANG[1], Qun YANG[2]

Computer translation is vital in daily life, and accuracy and fluency are key. To enhance machine translation quality, a bidirectional oblique scattering loss function is studied to optimize the training algorithm. A word - segmentation - based English machine translation tool, combining the byte pair encoding algorithm and segmentation metric, is proposed. Experiments show it has better semantic coherence, translation consistency, etc. With a 97.67% translation accuracy, 0.250s average latency, and shorter word recognition time, it better meets user needs. This research offers a new optimization method, improving machine translation performance.

Keywords: word slicing; loss function; machine translation; greedy search; column search

1. Introduction

With globalization and frequent international communication, English, an international language, is widely used in various fields [1]. However, non-native speakers face difficulties in understanding and using it. English machine translation, an effective tool, helps overcome language barriers [2]. In recent years, word segmentation algorithms in natural language processing have advanced, improving the accuracy and efficiency of English machine translation [3]. Big data and AI technology offer abundant data and computing power for better translation effects [4]. Existing English machine translation research has limitations. Algorithm accuracy affects translation results. Machine translation systems require improvement in recognizing and processing culturally rich vocabulary and expressions, which often convey nuanced cultural meanings [5]. Inadequate context handling leads to errors. Also, algorithms and models need continuous optimization. To solve these problems, a word segmentation algorithm - based English machine translation method is proposed. It optimizes existing word segmentation algorithms. The study combines forward and backward Kullback Leibler (KL) scattering for a new loss function and proposes a generalized unsupervised sub - word tangent framework by combining the byte - pair encoding algorithm with tangent measurement. It also combines Chinese

[1] Guangzhou Huashang College, Guangzhou 511300, China, e-mail: 86yangshuyue@163.com
[2] Guangzhou Huashang College, Guangzhou 511300, China, e-mail: yxxf828@163.com

word segmentation criteria with bytes and byte - pair encoding. The research contributes by proposing an English machine translation system based on word segmentation algorithm, optimizing the loss function and sub - word slicing algorithm in the Neural Machine Translation (NMT) system. The framework includes a review of relevant research, optimization of the loss function and word sorting algorithm, validation of new algorithms, and research conclusions. In the field of natural language processing, word segmentation algorithms are of great importance to languages without clear word boundaries, such as Chinese, Japanese and Thai. After optimization, the proposed English machine translation method based on word segmentation algorithm not only performs well in English translation, but also provides reference for machine translation in these languages. By improving and optimizing the word segmentation algorithm, the accuracy and efficiency of this kind of language machine translation can be effectively improved, and the language barrier can be overcome to help more extensive international communication.

2. Related work

Machine translation and neural network technology have made significant progress and have various applications in cross lingual communication and multilingual information processing. Shi C proposes a deep learning intelligent language translation model based on perceptual correlation cost function to amplify the impact of language translation using artificial intelligence, and the translation system of this model has excellent performance [6]. To enhance the impact of clever English translation, Song X proposed a model of English intelligent translation system based on multi-objective optimization algorithm. The results show that the model is quite sophisticated and can fulfil the demands of actual translation [7]. Yuan Z et al. proposed a fuzzy algorithm-based method for translating from English to improve the accuracy of English translation system, and the results showed that the technology may, to a certain degree, remove semantic ambiguity from the translation process and increase translation accuracy [8]. Sanjanasri J P et al. proposed a phrase-based statistical machine translation (PB-SMT) system for Indian languages to improve the performance of conditional probability and language models of phrase tables, and the results showed that the system outperforms the standard PB-SMT system [9]. Fahad A A et al. proposed a PB-SMT system for Indian languages based on deep learning to promote an intelligent translation system based on convolutional neural network with deep learning algorithm is proposed for effective communication between people with speaking disabilities and normal people. The overall precision of the proposed solution is 96.68%, which is appropriate and reliable for deaf people [10]. Li et al. adopted a solution combining edge computing and artificial intelligence, aiming to improve the performance of machine translation and user reading experience. By using a skip graph model to study user input features and applying a beetle antenna search neural model with attention long short-term memory to develop text sorting values, this method can identify edge servers and initiate corresponding translation processes to maximize translation matching. The experimental results indicate that this method can effectively improve the language comprehension ability of English learning and reduce learning difficulties [11].

Neural networks, as an artificial intelligence technique for distributed parallel information processing, extensively used in voice recognition, speech synthesis, and picture recognition. Cai L et al. propose a data quantization-based complex-valued memetic neural network control scheme to handle the issue of limited communication resources and fault interference to the system, and the model is effective and less conservative [12]. Chen Z proposes a Simulated Annealing-Hybrid Harmony Genetic Algorithm (SA-HHGA) optimization-based RBF to investigate the impact of using clever learning algorithms. The optimized Radial Basis Function (RBF) neural network's projected contextual values in 15 examples closely match the actual scenario values, demonstrating the network's strong predictive power [13]. Cheng et al. explored the application of deep learning technology in English oral pronunciation recognition and improved a multi-layer perceptron integrated with spike neural networks. The experimental results indicate that this method helps students identify differences between pronunciation and standards, correct pronunciation errors, and thus improve their English speaking performance [14].

Although previous research has made significant progress in machine translation and neural network technology, there are still some shortcomings and limitations. On the one hand, the translation quality of the existing model is difficult to guarantee, and the semantic accuracy and fluency are affected. On the other hand, the common loss function has some optimization space in the training process, and can not fully adapt to the complex and changeable translation scenarios. In addition, the traditional byte pair coding algorithm lacks effective new metrics in dealing with subword segmentation, which affects the further improvement of translation performance. Therefore, the proposed English machine translation method based on word slicing algorithm is studied. Different from the previous methods, this paper introduces a new segmentation metric to optimize the byte pair encoding algorithm, constructs a bidirectional skew divergence loss function, and adopts word frequency weighted segmentation metric to improve the existing problems from multiple dimensions, so as to improve the performance of machine translation.

3. Design of English machine translation based on word slicing algorithm

Recently, the use of computer translation in everyday life has become crucial, and how to obtain satisfactory accuracy and fluency is a key issue. This section focuses on the optimization measures of loss functions in NMT systems, while introducing a new cut metric optimized byte pair encoding algorithm to alleviate the problem of rare word representation.

3.1 Framework of neural network-based machine translation system

Common models in NMT include log-linear language models, neural network-based language models, and encoder-decoder models [15]. The log-linear language model is based around features to calculate the conditional probability of a particular word in a given context, and the probability of a word with a higher score is also higher, and the score vector is calculated as in equation (1).

$$s = Wx + b \tag{1}$$

In equation (1), W denotes the weight matrix, b is the bias vector, and x is the given contextual feature. The scores are normalized to the (0, 1) interval by applying the softmax function to output them in probabilistic form as in equation (2).

$$\begin{cases} l = soft\max(s) \\ l_j = \dfrac{\exp(s_j)}{\sum_{j'} \exp(s_{j'})} \end{cases} \tag{2}$$

In equation (2) l_j is the term j in the vector l. Gradient explosion and vanishing issues are common challenges for Recurrent Neural Networks RNNs in neural network-based language models. Thus, the study adopts Long short-term memory (LSTM), a RNN variant. LSTM has forgetting, input and output gates to manage memory unit states, better capturing long - range dependencies. The encoder-decoder model processes the source language information by first encoding it as a real-valued vector (i.e., hidden state) through a neural network, and then decoding the vector into a sentence in the target language through another neural network [16-17]. The hidden state of the target language is accepted and transformed into a probability at the softmax layer, which is calculated as in equation (3).

$$\begin{cases} m_t^f = embed^f(y_t^f) \\ h_t^f = \begin{cases} RNN^f(m_t^f, h_{t-1}^f), t \geq 1 \\ 0, \quad \quad otherwise \end{cases} \\ m_t^e = embed^e(g_{t-1}^e) \\ h_t^e = \begin{cases} RNN^e(m_t^e, h_{t-1}^e), t \geq 1 \\ h_{|F|}^f, \quad \quad otherwise \end{cases} \\ k_t^e = soft\max(W_{hs}h_t^e + b_s) \end{cases} \tag{3}$$

In equation (3) y is the source language sentence with the length of $|F|$, g is the target language sentence with the length of $|E|$, t is the time step, m_t^f is the word vector, $RNN^f(\cdot)$ is the encoder, $RNN^e(\cdot)$ is the decoder, h_t^f is the encoder hidden state of the source language sequence, h_t^e is the decoding state, $h_{|F|}^f$ is the final state of the source language encoding, and k_t^e is the probability. The encoder-decoder model for machine translation has issues with long-distance dependencies and encoding long sentences. The attention mechanism is introduced to solve them. The studied NMT system uses an encoder-decoder with attention and LSTM units, as in Fig. 1.

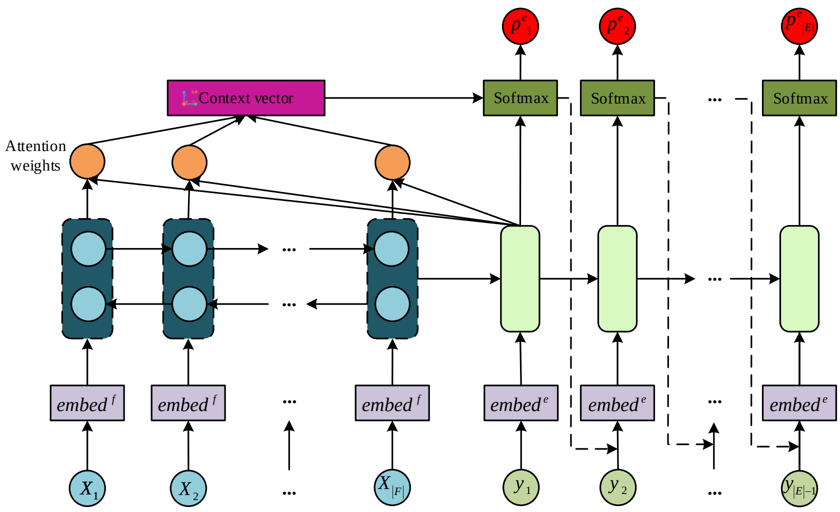


Fig. 1. The architecture of baseline model

After calculating conditional probabilities, search algorithms are used for translation. Common ones in NMT are greedy and beam search. Greedy picks the best at each stage but may miss the global optimum. Beam keeps the top-n best-translated words at each step. To solve this problem, when using column search, the model divides conditional probability by target length to generate shorter sentences. NMT's parameter optimization combines Adam and stochastic gradient descent (SGD) in small batches, and performance is evaluated by Bilingual Evaluation Understudy (BLEU) as in equation (4).

$$BLEU = BP \cdot \exp(\sum_{n=1}^{N} w_n \log z_n) \tag{4}$$

In equation (4), z_n is the corresponding n-gram accuracy score, w_n is the weight of each score, and BP is the length penalty factor.

3.2 Optimal design of loss function based on neural network machine translation

During the NMT training, the study constructs a neural network language model to calculate the probability of the target utterance and uses a cross-entropy loss function based on the maximum likelihood (ML) criterion to obtain the optimal value of the parameters [18]. The cross-entropy loss function has the advantages of fast convergence and avoiding gradient disappearance, and is represented by the word-level cross-entropy at each time step of its decoding in the practical calculation, and the vector form of the cross-entropy loss is as in equation (5).

$$J_{XENT} = -\sum_{t=1}^{n} f_t \log(\hat{f}_t) \tag{5}$$

In equation (5), f_t is a one-hot vector corresponding to the correct target label of the time step t, and $\hat{f}_t$ is the probability vector predicted by the model, both of which have dimensions equal to the word list size. KL scatter is used to calculate the separation of two randomly distributed probability distributions, which is represented in equation (6).

$$D_{KL}(P \| Q) = E_{x \sim p}[\log P(x) - \log Q(x)] \tag{6}$$

In equation (6), P is distribution of the target language's actual data and Q is the distribution predicted by the model. The connection between KL scatter $D_{KL}(P \| Q)$ and cross-entropy $H(P,Q)$ is represented in equation (7).

$$H(P,Q) = H(P) + D_{KL}(P\|Q) \tag{7}$$

In equation (7), $H(P)$ is the entropy of the data itself, which is independent of Q and is equivalent to a constant, so the relationship between KL scatter and cross-entropy is as in equation (8).

$$H(P,Q) = D_{KL}(P\|Q) \tag{8}$$

Swapping the positions of P and Q yields the reverse form of $D_{KL}(Q\|P)$, as in equation (9).

$$D_{KL}(Q\|P) = E_{x\sim Q}[\log Q(x) - \log P(x)] \tag{9}$$

In NMT, there are often multiple optimal choices at the same time, and there are gains and losses when using the word level $D_{KL}(P\|Q)$ and $D_{KL}(Q\|P)$ as training targets. Therefore, to solve the problem that $D_{KL}(Q\|P)$ is not fully defined at $P=0$, the study uses $\alpha-$ skew divergence ($\alpha-$ skew divergence), which is expressed in equation (10).

$$s_\alpha(Q,P) = D_{KL}(Q\|\alpha Q + (1-\alpha)P) \tag{10}$$

In equation (10), α is the degree of approximation of the function to $D_{KL}(Q\|P)$, which takes the value range [0, 1]. When $\alpha = 0.01$ is used, $\alpha-$ skew scatter is the best way to simulate the process of minimizing $D_{KL}(Q\|P)$. Similarly, $s_\alpha(P,Q)$ is used to simulate the process of minimizing $D_{KL}(P\|Q)$. The new loss function is obtained by further linear interpolation in both directions of the $\alpha-$ skewed scatter, called the bidirectional skewed scatter (DSD), which is represented in equation (11).

$$D_{DS} = \beta s_\alpha(P,Q) + (1-\beta)s_\alpha(Q,P) \tag{11}$$

In Eq. (11), β is the interpolation coefficient, which takes values in the range [0, 1]. The word levels $D_{KL}(P\|Q)$ and $D_{KL}(Q\|P)$ are obtained, then the two objectives are approximated using the skewed scatter of $\alpha-$, and after the approximation, linear interpolation is performed to obtain the final expansion of the new loss function. In summary, the loss function optimization of the DSD-NMT model is completed.

3.3 Optimal design of word slicing algorithm in neural network machine translation

In machine translation, <UNK> tokens for out-of-vocabulary words limit performance. This study extends NMT's byte-pair encoding to a generalized unsupervised subword slice framework with new metrics, a combined Byte Pair Encoding (BPE)-based algorithmic framework as in Fig. 2.

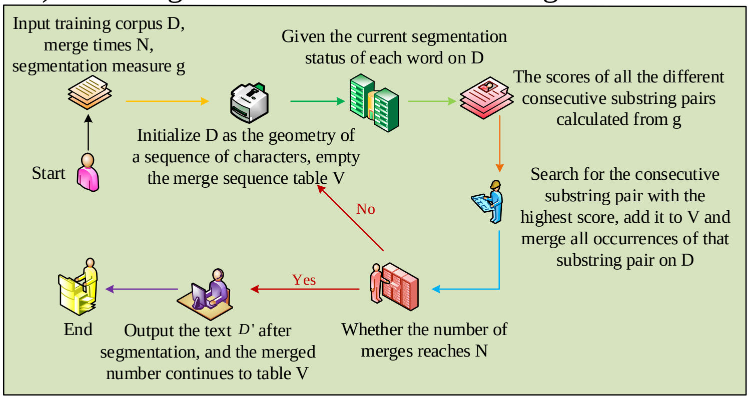


Fig. 2. Algorithm steps of generalized byte pair encoding subword segmentation

As shown in Fig. 2, the cut-off metric impacts a subword's comprehensibility. The study analyzed four cut-off schemes: for substring frequency (FRQ), adjacency variety (AV), description length gains (DLG), and word frequency weighting. FRQ counts substring occurrences, with candidates having frequency>1. AV evaluates if a substring can be an independent word, defined in equation (12).

$$g_{AV}(s_i) = AV(s_i) = \min\{L_{av}(s_i), R_{av}(s_i)\} \tag{12}$$

In equation (12), s_i is the subword, $L_{av}(s_i)$ is the number of types of different antecedent tokens of s_i, and $R_{av}(s_i)$ denotes the variety of s_i's successor tokens in existence. DLG, as a measure of lexical richness, is often used in word boundary prediction tasks. shannon-Fano encoding length is the description length of a piece of text and is expressed as in equation (13).

$$DL(X) = -\sum_{X \in V_X} c(x) \log \frac{c(x)}{|X|} \tag{13}$$

In equation (13) X is the text, V_X is the list of words containing all tokens in X, $c(x)$ is the count of tokens x in X, and $|X|$ is the total number of tokens in X. For a particular substring of in X s_i, the change in the length of the text description when replacing s_i with the index r is DLG, and the scoring function is as in equation (14).

$$g_{DLG}(s_i) = DLG(s_i) = DL(X) - DL(X[r \to s_i] \oplus s_i) \tag{14}$$

In equation (14), $X[r \to s_i]$ is the new text obtained after replacing all s_i in X with r, and $\oplus$ is a separator splice between two strings. Most of the words in the NMT model are only low-frequency words rather than real words, and these low-frequency words interfere with the calculation of FRQ and AV, so the study introduces a word-frequency weighted cut-off metric, which is represented in equation (15).

$$\begin{cases} g'_{FRQ}(s_i) = g_{FRQ}(s_i) \sum\limits_{\forall w, s_i \in w} f(w) \\ g'_{AV}(s_i) = g_{AV}(s_i) \sum\limits_{\forall w, s_i \in w} f(w) \\ g'_{DLG}(s_i) = g_{DLG}(s_i) \sum\limits_{\forall w, s_i \in w} f(w) \end{cases} \tag{15}$$

In equation (15), $f(w)$ denotes the frequency of occurrence of the word w in the corpus. In summary, the optimization of the subword cutting algorithm in NMT is completed.

In summary, in the design of machine translation system, the log-linear language model, neural network language model and encoder-decoder model are adopted. The LSTM unit is used to solve the gradient problem of the recurrent neural network, and the attention mechanism is introduced to solve the encoder-decoder model to deal with long distance dependence and long sentence coding. In the training process, the neural network language model is constructed, and the parameters are optimized by using the cross entropy loss function based on the maximum likelihood criterion, and then the DSD loss function is proposed for further optimization. At the same time, to solve the problem that unknown words affect the performance of machine translation, byte pair coding is extended to a generalized unsupervised subword slicing framework, and four segmentation schemes such as substring frequency, adjacency diversity, description length gain and word frequency weighting are analyzed to optimize the subword slicing algorithm, so as to improve the accuracy and fluency of English machine translation.

4. Performance analysis of English machine translation based on word slicing algorithm

To test the efficiency and viability of the study's suggested word syncopation algorithm-based English machine translation, this section focuses on testing the effects of interpolation coefficient β, column search width, and syncopation metric on the translation performance, and further tests the convergence curves of semantic coherence, translation consistency and fluency as well as the correctness rate and time delay of translation results of the system.

4.1 Performance analysis of neural network-based machine translation with optimized loss function

To verify the optimization of the neural network - based loss function for machine translation, the study uses IWSLT2015 dataset. It evaluates translation quality with 4-gram BLEU. Four language-pair training sets are processed with tokenization and joint byte pair encoding. The study uses greedy search to test the effect of different interpolation coefficients β on the translation quality in four training sets.

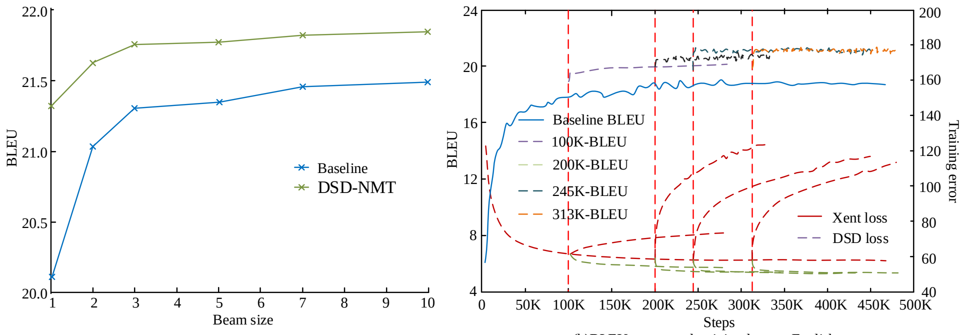
Table 1

BLEU scores with different β values are used on different test sets

Test set	English-German	English-French	English-Chinese	English - Thai
Baseline	20.11	20.11	20.11	20.11
$\beta=1$	20.41	20.32	20.15	20.76
$\beta=0.8$	20.62	20.56	20.31	20.96
$\beta=0.6$	20.83	20.75	20.47	21.15
$\beta=0.4$	21.05	20.97	20.62	21.35
$\beta=0.2$	21.26	21.18	20.78	21.54
$\beta=0$	21.32	21.40	20.94	21.74

Table 1 shows the BLEU scores when training with different β values on the four test sets. The BLEU scores of all four test sets are highest when $\beta=0$ is used, i.e., the model trained with $\beta=0$ has the best translation quality, so $\beta=0$ is taken for all subsequent tests. To investigate the effects of column search width and loss function switching on translation quality, the study tested the variance in

BLEU scores on the English-French test set between the Institute's planned DSD-NMT and NMT baseline systems with different column search widths, and also tested the cross-entropy loss of the DSD-NMT model when switching loss functions.



(a)BLEU scores with different beam sizes on English-French dev set

(b)BLEU scores and training loss on English-French dev set with DSD switchin

Fig. 3. Effect of column search width and loss function switching on translation quality

Fig. 3(a) shows BLEU scores of DSD-NMT and NMT baselines with varying column search widths. The best scores are at 10, and the difference shrinks as the width increases. Fig. 3(b) presents BLEU scores and training errors in the English-French task. Switching to DSD loss boosts BLEU scores, with the best point near the cross-entropy loss convergence. The cross-entropy loss rises after the switch, suggesting the model may reach a local optimum, proving the loss-function switch helps the model globally.

4.2 Performance analysis of optimized subword slicing algorithm in NMT

To verify the optimized word slicing algorithm, experiments on the English-French test set were done. Set word & attention vec dims to 256, lr to 0.001, batch size to 32, train 40 epochs, halve lr every 10 epochs, and count BLEU scores.

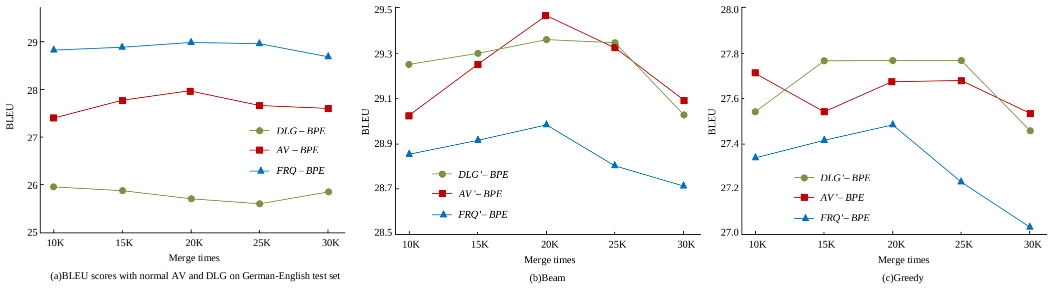


Fig. 4. BLEU scores for the general segmentation metric and frequency weighting scheme segmentation metric on the English-French test set

Fig. 4(a) shows the translation performance of FRQ-BPE, AV-BPE, and DLG-BPE in the English-French task. FRQ-BPE and AV-BPE outperform DLG-BPE, with FRQ-BPE being the best. Figs. 4(b) and 4(c) show BLEU scores. For greedy search, DLG'-BPE and AV'-BPE perform better. At *N=20K* in column search, they improve BLEU by 0.47 and 0.374 points over FRQ'-BPE.

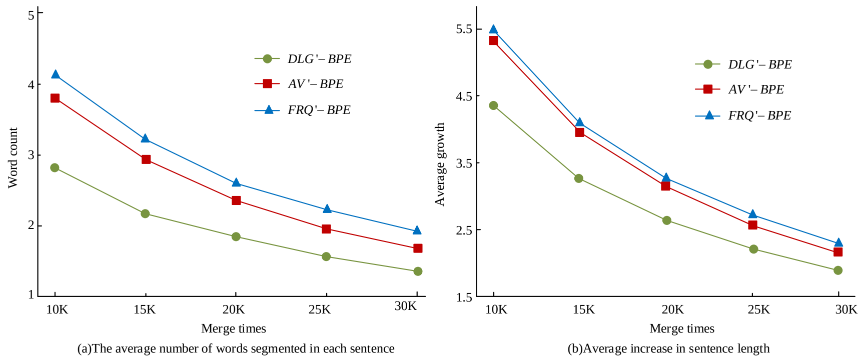


Fig. 5. Statistics of words and sentence after segmentation

Fig. 5(a) shows DLG'-BPE and AV'-BPE have more complete word forms per sentence than FRQ'-BPE at the same merge number. Fig. 5(b) shows they better control sequence length, with AV'-BPE's best performance between the curves.

4.3 Comprehensive performance analysis of English machine translation based on word slicing algorithm

To test the comprehensive performance of the proposed word-sorting algorithm-based machine translation model, the study is trained on the English-French test set for convergence. The proposed DSD-NMT model is also compared with the Minimum risk training (MRT) model, Bahdanau-log-likelihood (Bahdanau-LL) model, Bahdanau-log likelihood-actor critic (Bahdanau-LL) model, and Bahdanau-log likelihood-actor critic (Bahdanau-LL) model [19]. Bahdanau-log likelihood-actor critic, Bahdanau-LL-AC) model were compared to obtain convergence curves.

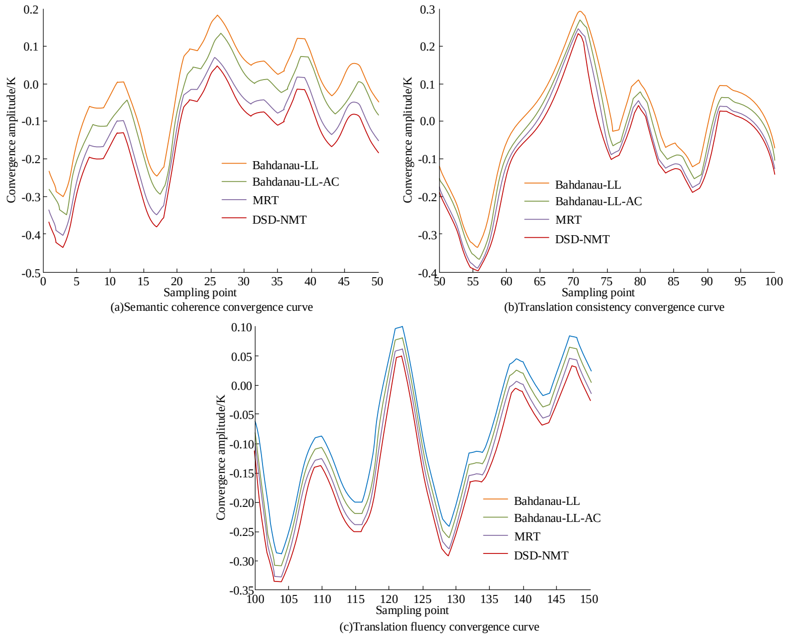


Fig. 6. Automatic conversion translation output convergence curve comparison

Figs. 6(a)-6(c) compare convergence curves of semantic coherence, translation consistency and fluency for four models. The DSD-NMT model has better translation accuracy and curves. Trained on four test sets, it's compared with Statistical Machine Translation (SMT) and NMT single models by BLEU scores.

Table 2

Performance of four test sets

Model	Method	English-German	English-French	English-Chinese	English - Thai
MRT	MRT+beam	20.45	34.23	20.90	25.61
Bahdanau-LL	ML+greedy	20.35	29.33	19.34	24.35
	ML+beam	20.40	30.71	19.92	24.13
Bahdanau-AC-LL	ML+AC+greedy	19.56	30.85	20.25	24.68
	ML+AC+beam	19.73	31.13	20.67	24.96
Baseline-SMT	MERT+greedy	18.83	31.55	19.35	23.49
	MERRT+beam	19.91	33.82	20.23	24.01
Baseline-NMT	ML+greedy	20.89	32.10	21.26	24.76
	ML+beam	22.13	34.70	22.41	24.79
DSD-NMT	DSD+greedy	22.02	33.56	23.45	25.98
	DSD+beam	22.60	35.04	23.89	26.13

Table 2 shows BLEU scores of four models, SMT, and NMT single models on four test sets. DSD-NMT has the best BLEU scores in greedy and column search. For English-German task, in greedy search, it outperforms SMT by 3.19 and NMT by 1.13 BLEUs. In column search, by 2.69 and 0.47 BLEUs. In other tasks, it also outperforms. Tests on 1000 utterances verify operational efficiency.

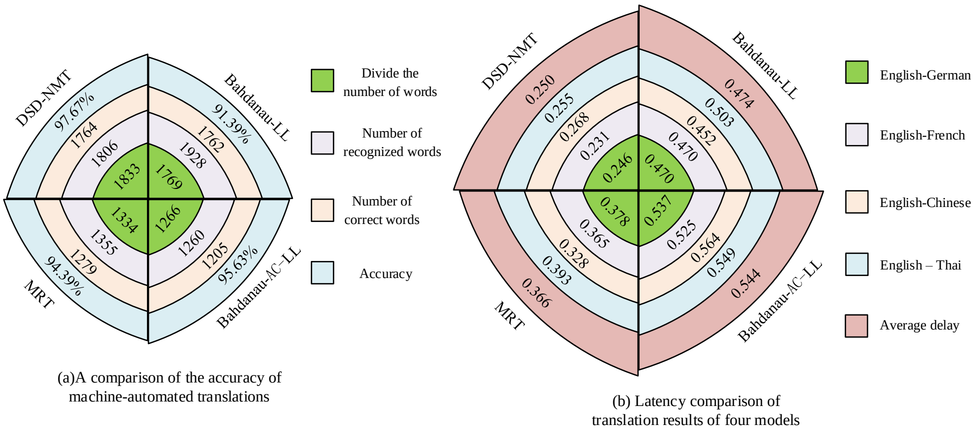


Fig. 7. Comparison of the accuracy of the automatic machine conversion translation and the delay of the translation result

Fig. 7(a) shows that DSD-NMT has the highest automatic machine translation accuracy at 97.67%. Fig. 7(b) shows that its average latency is 0.250s, which is lower than that of MRT (0.366s), Bahdanau-AC-LL (0.544s) and Bahdanau-LL (0.474s), and the computational complexity is reduced by 45.7%. DSD-NMT performs well in terms of translation quality, latency and cost effectiveness, providing an effective machine translation solution.

5. Conclusions

Machine translation technology is becoming more and more important in political, economic and cultural exchanges. However, at present, there are some problems such as the defects of training algorithms and the difficulty of expressing rare words. This paper proposes an English machine translation system based on word segmentation algorithm, namely DSD-NMT model. The experimental results show that the performance of BLEU peak mean value and DLG'-BPE and AV'-BPE at *N=20K* is better than FRQ'-BPE when word frequency weighted cut score is applied. Compared with MRT, Bahdanau ACLL and Bahdanau LL models, DSD-NMT model has better curve convergence in terms of semantic coherence, translation consistency and fluency, higher translation quality, and a translation accuracy of 97.67%. They are 3.28%, 2.04%

and 6.28% higher than the above models respectively, and the word recognition time is only 0.250s, which can better meet the needs of users. Although this research method has limitations, it can provide reference for subsequent research, and can be applied to multi-domain and low-resource languages in the future, to build a better machine translation system by optimizing algorithms, exploring domain adaptation, etc., and to collect user data feedback to provide more personalized services.

Abbreviation List

Full name	Abbreviation
Kullback Leibler	KL
Neural Machine Translation	NMT
phrase-based statistical machine translation	PB-SMT
frames per second	FPS
full high definition	FHD
Simulated Annealing-Hybrid Harmony Genetic Algorithm	SA-HHGA
Radial Basis Function	RBF
Recurrent Neural Networks	RNNs
Long short-term memory	LSTM
stochastic gradient descent	SGD
small batches. descent	SGD
Bilingual Evaluation Understudy	BLEU
maximum likelihood	ML
bidirectional skewed scatter	DSD
Byte Pair Encoding	BPE
substring frequency	FRQ
adjacency variety	AV
description length gains	DLG
Minimum risk training	MRT
Bahdanau-log-likelihood	Bahdanau-LL
Statistical Machine Translation	SMT

REFERENCES

[1] *Liang W, Xu X.* Savitar: an intelligent sign language translation approach for deafness and dysphonia in the COVID-19 era. Data Technologies and Applications, 2024, 58(2): 153-175.

[2] *Ting F K, Tan K H.* Enhancing English Language Vocabulary Learning among Indigenous Learners through Google Translate. Journal of Education and E- Learning Research, 2021, 8(2): 143-148.

[3] *Chimalamarri S, Sitaram*. Linguistically enhanced word segmentation for better neural machine translation of low resource agglutinative languages. International Journal of Speech Technology, 2021, 24(4), 1047-1053.

[4] *Abidin Z, Permata P, Ariyani F.* Translation of the lampung language text dialect of nyo into the Indonesian Language with DMT and SMT approach. INTENSIF: Jurnal Ilmiah Penelitian Dan Penerapan Teknologi Sistem Informasi, 2021, 5(1): 58-71.

[5] *Popova N V, Stepanova M M, Kuzmina A V*. Methodological aspects of the application of audio-visual tools for learning a foreign language and translation at the university. Vestnik of Samara State Technical University Psychological and Pedagogical Sciences, 2021, 18(1): 87-98.
[6] *Shi C.* Research on intelligent language translation system based on deep learning algorithm. Soft Computing, 2022, 26(16): 7509-7518.
[7] *Song X*. Intelligent English translation system based on evolutionary multi-objective optimization algorithm. Journal of Intelligent & Fuzzy Systems, 2021, 40(4): 6327-6337.
[8] *Yuan Z, Jin C, Chen Z.* Research on language analysis of English translation system based on fuzzy algorithm. Journal of Intelligent & Fuzzy Systems, 2021, 40(4): 6039-6047.
[9] *Sanjanasri J P, Kumar M A, Soman K P*. Deep learning-based techniques to enhance the precision of phrase-based statistical machine translation system. International Journal of Computer Aided Engineering and Technology, 2020, 13(1-2): 239-257.
[10] *Fahad A A, Hassan H J, Abdullah S H.* Deep Learning-based Deaf & Mute Gesture Translation System. International Journal of Science and Research (IJSR), 2020, 9(5): 288-292.
[11] *Li S, Huang Y.* BAS-ALSTM: analyzing the efficiency of artificial intelligence-based English translation system. Journal of Ambient Intelligence and Humanized Computing, 2024, 15(1): 765-777.
[12] *Cai L, Xiong L, Cao J, Zhang H, Alsaadi F.* State quantized sampled-data control design for complex-valued memristive neural networks. journal of the Franklin Institute, 2022, 359(9): 4019-4053.
[13] *Chen Z.* Research on internet security situation awareness prediction technology based on improved RBF neural network algorithm. Journal of Computational and Cognitive Engineering, 2022, 1(3): 103-108.
[14] *Cheng S, Yang Q, Luo H.* Design of Neural Network-based Online Teaching Interactive System in the Context of Multimedia-assisted Teaching. Informatica, 2024, 48(7) 53-62.
[15] *Yan J, Jin L, Yuan Z, Liu Z.* RNN for receding horizon control of redundant robot manipulators. IEEE Transactions on Industrial Electronics, 2021, 69(2): 1608-1619.
[16] *Kesiman M, Dermawan K T.* AKSALont: Automatic transliteration application for Balinese palm leaf manuscripts with LSTM Model. Journal Teknologi dan Sistem Komputer, 2021, 9(3): 142-149.
[17] *Sun W, Zhang G, Zhang X, Zhang X, Ge N*. Fine-grained vehicle type classification using lightweight convolutional neural network with feature optimization and joint learning strategy. Multimedia Tools and Applications, 2021, 80(20), 30803-30816.
[18] *Lin S C, Lin J*. A dense representation framework for lexical and semantic matching. ACM Transactions on Information Systems, 2023, 41(4), 1-29.
[19] *Thorat S A, Jadhav K P*. Improving conversation modelling using attention based variational hierarchical RNN. International Journal of Computing, 2021, 20(1), 39-45.