

Received 25 June 2024, accepted 6 July 2024, date of publication 11 July 2024, date of current version 22 July 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3426652

RESEARCH ARTICLE

Optimized English Translation System Using Multi-Level Semantic Extraction and Text Matching

HUI YANG 

School of Foreign Languages, Anhui Agricultural University, Hefei, Anhui 230036, China

e-mail: yhahau123456@126.com


This work was supported by the 2022 Scientific Research Projects of Colleges and Universities in Anhui Province (Philosophy and Social Sciences) "A Study on the Translation Activities of Li Jiye, an Anhui Scholar—Based on Bourdieu's Sociological Perspective" under Project 2022AH050857.

ABSTRACT The domain of machine text translation and matching is undergoing substantial transformations amidst the perpetual evolution of deep learning methodologies. By amalgamating the contemporary realm of generative models and networks with the multi-faceted attentiveness of multiple heads, there has been a pronounced enhancement in the efficacy of existing text translation and matching endeavors. Consequently, this manuscript endeavors to elucidate the intricacies of the text-matching conundrum within the ambit of English translation. It posits a novel MA-Transformer text-matching framework that seamlessly integrates multi-tiered semantic feature extraction methodologies to actualize the text-matching task in the English translation process. The framework initiates its journey by employing Continuous Bag of Words (CBOW) for word vector embedding, thereby accomplishing the generation and embedding of word vectors. Subsequently, it expeditiously conducts the multilevel amalgamation of data features through the expeditious execution of the multi-head Transformer model. Following the culmination of feature fusion, a judicious sequence of data downgrading and feature screening ensues, ultimately culminating in the attainment of high-precision text matching. The experimental results show that the constructed MA Transformer model performs well in public and actual data testing, with an average precision of 0.867 and 0.722, respectively, on the two types of datasets. The accuracy of the text-matching is higher than that of the current common method frameworks, which provide technical support and references for the future construction of English translation systems.

INDEX TERMS Multi-level semantic feature extraction, machine translation, English translation, text matching.

I. INTRODUCTION

In light of the escalating global exchange of information, an imperative arises for precise matching and translation across multilingual texts. Conventional translation systems, constrained by inherent limitations in handling semantic intricacies and contextual nuances, have prompted researchers to delve into the extraction of information from multi-tiered semantic perspectives through the prism of deep learning techniques. This concerted effort aims at augmenting the

The associate editor coordinating the review of this manuscript and approving it for publication was Muhammad Asif .

efficacy of translation systems in response to the exigencies of our interconnected world. In the current epoch dominated by the relentless advancement of Artificial Intelligence (AI), machine translation (NMT) and pre-trained language models stand as the focal points of investigation within this domain [1]. Neural Machine Translation (NMT) harnesses deep neural networks to achieve seamless end-to-end translation, exemplified prominently by the Transformer model. This model is celebrated for its prowess in capturing contextual information and linguistic structure via the sophisticated self-attention mechanism. Simultaneously, the integration of pre-trained language models like BERT and GPT enriches the

contextual reservoir for the translation task, thereby elevating the overall quality of translation [2]. The synergistic application of these methodologies propels the continual evolution of AI translation, laying a robust foundation for the realization of more precise and eloquent multilingual translations.

In recent epochs, the strides witnessed in text matching and translation systems, underpinned by the bedrock of deep learning, have been nothing short of extraordinary. Amidst this paradigm shift, the exploration of multi-level semantic feature extraction has emerged as a focal point of noteworthy research endeavors. Notably, scholars have attained new heights by extracting multilevel semantic information from input text through the infusion of pre-trained language models endowed with attention mechanisms and hierarchical structures [3]. This includes more than just the subtle meanings of words, extending to cover relationships within sentences and paragraphs, thus encompassing inter-textual connections more thoroughly. Moreover, the scholarly community has directed its attention towards cross-lingual translation scenarios, endeavoring to integrate the extraction and alignment of multilevel semantic information into translation systems for a discernible enhancement in translation quality. These methodologies encompass sophisticated techniques such as semantic space mapping and multilingual representation learning, ensuring a more adept retention and seamless transfer of semantic information across diverse linguistic domains [4].

Within the realm of text translation and matching research, pivotal advancements are attributed to the integration of deep learning models, with Transformer, BERT, and GPT emerging as pivotal technologies [5]. The Transformer model, leveraging the self-attention mechanism, adeptly addresses long-distance dependencies, thereby demonstrating prowess in translation tasks. BERT, grounded in a pre-trained language model, fortifies text comprehension and excelling in contextual understanding and feature extraction, making substantial contributions to text matching. Conversely, GPT adopts a generative architecture, showcasing commendable performance in generative and translation tasks, particularly in articulating intricate textual nuances. In parallel, LSTM and CNN bring forth distinctive advantages in processing sequential and local features. These deep learning models, characterized by their capability to capture intricate semantic relationships and enhance contextual comprehension, play a pivotal role in significantly elevating performance in text translation and matching tasks [6]. Hence, the application of deep learning methodologies in resolving the text-matching intricacies within the context of English translation holds the potential to markedly enhance model recognition efficacy and alleviate the workload for pertinent personnel. In the paper, existing neural network models are combined to enrich the content of feature extraction and achieve high-precision text matching. At the same time, this model is innovatively used to test actual data and further strengthen its application scenarios. The specific contributions of this paper are as follows:

1. In addressing the translation matching challenge within the ambit of English translation, a text recognition framework predicated on the Transformer model is postulated to consummate the alignment of disparate textual contents.

2. Employing word vector embedding technology and feature fusion techniques for the refinement of the Transformer model, the MA-Transformer model is meticulously crafted to realize high-precision text matching.

3. By subjecting the model to testing using both existing datasets and a bespoke English translation dataset, the objective is to achieve elevated precision in text matching across diverse datasets. The outcomes elucidate that the efficacy of the MA-Transformer method surpasses that of conventional text matching models.

The rest of the paper is organized as follows: Section II introduces the related works for machine translation and text matching. The proposed framework is established in Section III. In section IV, the experiment details and results are given, and the conclusion is drawn at the end.

II. RELATED WORKS

A. NEURAL NETWORK MACHINE TRANSLATION RESEARCH

Since the emergence of neural network-based machine translation, its remarkable performance has captivated the attention of researchers, prompting a convergence of traditional methodologies like statistical machine translation with neural approaches, resulting in remarkable translation outcomes. This integration has catapulted neural networks into the forefront of contemporary language modeling, with Schwenk's neural network-based language model standing out as a seminal contribution to the trajectory of neural network-based modeling research [7]. The inception of neural machine translation models traces back to Forcada's 1997 proposal [8]. Hindered by the scarcity of large corpora during that era and the absence of the expansive datasets available today, progress in this domain remained subdued for an extended period. However, with recent advancements in neural network technology and the continual augmentation of artificial intelligence computational capabilities, neural network methodologies in machine translation research and application have experienced a gradual and expansive evolution. Sutskever's introduction of a neural network machine learning model with an encoding and decoding structure in 2016 marked a significant milestone [9]. Subsequent breakthroughs, such as Junczys-Dowmunt's impactful demonstration of the superiority of neural machine translation over traditional methods in 15 language pairs and 30 translation directions, formally heralded the era of neural machine translation [Wu et al. integrated the concept of residual connection into the neural machine translation model, aiming to address the challenge of gradient disappearance by preserving input information directly in the output [11]. Gehring et al. advanced a convolutional neural network-based neural machine translation model, surpassing Google's machine translation model in both accuracy and translation speed [12].

Further innovations emerged as Sennrich et al. introduced a seq2seq model with two models operating in opposite directions from the source language to the target language and vice versa. This holistic model effectively translated monolingual data [13]. Ren et al. introduced the Triangular Architecture Neural Machine Translation (TANMT) model, leveraging the abundant alignment corpus of a major language to enhance machine translation capabilities for a smaller language through the incorporation of a language with a rich alignment corpus, forming a triangular structure [14].

B. TEXT MATCHING RESEARCH

Text matching, as a quintessential task in the realm of natural language processing, has undergone continuous evolution and finds widespread applications in both production and daily life. In its nascent stages, literal matching served as the foundation for text similarity calculation, quantifying similarity based on the ratio of word matches between two texts to the total text length. The introduction of TF-IDF considered the importance of word frequency and inverse document frequency, extracting keywords to mitigate the impact of generic words on text similarity calculation [15]. Building upon TF-IDF, BM25 introduced adjustable parameters to enhance overall flexibility in similarity computation [16]. Subsequent advancements witnessed the introduction of machine learning-based approaches, including text similarity computation grounded in Latent Semantic Analysis (LSA) [17] and Latent Dirichlet Allocation (LDA) [18]. LDA, a topic model, considers the overall information of the text in similarity calculations, broadening the scope of observation. Contemporary research on text matching predominantly centers around the design of neural network models, categorized into semantic representation-based models and semantic interaction-based models. Semantic representation-based models employ neural networks to independently learn distributed representations of sentence pairs, utilizing classifiers for binary classification tasks or cosine similarity for text matching degrees. Examples include DSSM (Deep Semantic Structured Model) [19], C-DSSM [20] with convolutional neural networks for enhanced feature extraction, and R-DSSM [21] utilizing recurrent neural networks tailored for temporal data to capture contextual semantic information.

In contrast, semantic interaction-based models address the challenge of independent text encoding, adopting match aggregation frameworks to align low-level information in two texts. DecomAtt [22] leverages an attentional mechanism to interact with semantic information between texts, utilizing a feed-forward network for information aggregation. ESIM [23], employing bidirectional LSTM, encodes text and incorporates an attentional mechanism to achieve semantic interaction. BiMPM introduces an advanced multi-view matching operation, extracting diverse interaction features across different horizons [24]. For the models used in the text matching research mentioned above, the author has

summarized their advantages and disadvantages, and the results are shown in Table 1:

TABLE 1. The characteristic for the current research.

Approach	Strengths	Weaknesses
Literal Matching	Simple implementation, straightforward	Ignores semantic information, limited to surface similarity
TF-IDF	Considers term frequency and document frequency, reduces impact of common words	May not capture deep semantic meaning, sensitive to exact wording
BM25	Adjustable parameters, higher flexibility	Complex parameter tuning, computationally intensive
LSA	Captures latent semantic relationships, dimensionality reduction	May not capture topic distributions well, computationally expensive
LDA	Considers overall document themes, probabilistic model	Requires large corpus for accurate theme modeling, complex
DSSM	Projects sentences into a common latent space, effective similarity prediction	Complex model, requires significant training data
C-DSSM	Strong feature extraction with CNNs	High computational cost, requires large datasets
R-DSSM	Captures sequential dependencies and context with RNNs	Computationally intensive, requires large datasets
DecomAtt	Uses attention mechanism for semantic interaction	Complex implementation, requires significant computational resources
ESIM	Bi-directional LSTM encoding with attention mechanism	High computational cost, complex implementation
BiMPM	Advanced multi-perspective matching for diverse interaction features	Complexity and computational cost, may require extensive tuning

The comprehensive overview presented above underscores the expansive application landscape of neural network methodologies within the domain of machine translation. Building upon the foundations laid by traditional machine translation, the continuous evolution of neural networks and word vector technologies has ushered in a new era of enhanced model analysis. This progress enables a more profound exploration of models by leveraging additional dimensions, thereby fostering a more nuanced examination of the text matching challenges inherent in machine translation. The amalgamation of existing word vector models and deep network architectures results in the creation of a text analysis model endowed with heightened data processing capabilities. This model not only facilitates text-matching but also delves into sentiment theme analysis. This synergistic approach not only bolsters the interpretability of semantic ambiguity during large-scale translation processes but also upholds the translation quality, thereby ensuring a more refined and nuanced analysis of machine translation challenges.

III. METHODOLOGY

A. WORD EMBEDDING METHODS IN Word2Vec

Word2Vec constitutes a category of models employed for representing words as vectors within a continuous vector space,

standing as a crucial technique in the domain of Natural Language Processing (NLP). At its core, this model operates on the fundamental principle of mapping each word onto a vector within a high-dimensional space. This mapping is achieved through the analysis of extensive text corpora which facilitates the learning of semantic relationships between words. The resulting vector representation brings words with similar semantics closer to the vector space, thereby enhancing the model's capacity to capture semantic information between words effectively. The Word2Vec model manifests itself through two primary architectures: CBOW and the Skip-Gram [25]. Its specific structure is shown in Figure 1:

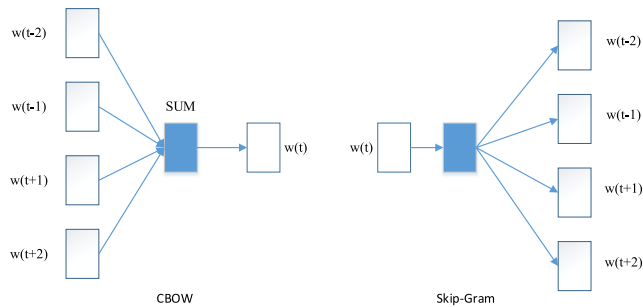


FIGURE 1. The Word2Vector model.

$w(t)$ is the center word. $w(t - 2)$, $w(t - 1)$, $w(t + 1)$ and $w(t + 2)$ are the contexts, and the window size in the figure is 5, which can be adjusted according to the task requirements. The primary objective of the CBOW model is to anticipate the current word based on the contextual words in its vicinity. To accomplish this, the model aims to predict the target word by considering the average of the surrounding context words within a specified window size. This process is mathematically represented by the formula shown in equation (1):

$$L = \sum_{w=V} \log p(w | \text{Context}(w)) \quad (1)$$

In the equation, V represents the corpus, and w stands for any word within the corpus of V . The primary training objective of the CBOW model is to minimize the prediction error, ensuring precise anticipation of the target word. The input to the CBOW model comprises the One-Hot encoding of context words within the specified window. Simultaneously, the hidden layer involves One-Hot encoding capable of mapping the input, typically of small dimensionality. Conversely, the output layer predicts the One-Hot encoding of the target word by leveraging the average of the context words [26].

The Skip-Gram model, in contrast to CBOW, aims to predict the surrounding words in the context of the current word. Specifically, given a target word, the model tries to predict the context words that may occur within a given window. Its computational procedure is shown in equation (2):

$$L = \sum_{w=V} \log p(\text{Context}(w) | w) \quad (2)$$

The Skip-Gram model, akin to the CBOW model, is trained through the minimization of prediction errors, ensuring the

accurate prediction of contextual words. However, for the specific needs outlined in this paper, the CBOW method has been selected for embedding and analyzing word vectors.

B. TRANSFORMER-BASED ENCODING INTERACTION

The Transformer represents a deep learning model founded on a self-attention mechanism specifically designed for processing sequential data, particularly within the domain of natural language processing. In the realm of natural language processing, conventional coding networks include recurrent neural networks and convolutional neural networks. However, both exhibit inherent limitations; RNNs are susceptible to gradient disappearance and have a diminished capacity for capturing long-distance dependencies. In response to these shortcomings, the Transformer network is introduced, aiming to fortify and enhance the model [27].

Before delving into the Transformer network, a brief elucidation of the attention mechanism is warranted. Attention mechanisms in text can be categorized into Soft Attention and Hard Attention. Soft attention entails allocating attention to all data points and assigning corresponding attention weights. In contrast, Hard Attention involves setting filtering conditions, resulting in certain attention weights being zero. Notably, this paper exclusively introduces soft attention mechanisms.

At the heart of the attention mechanism lies the computation of attention scores between two vectors, often referred to as alignment scores. Suppose the vector $u \in R^d$ are request vectors, and the object of the attention mechanism is $V = \{v_i\} \in R^{n*d}$ is a vector consisting of n a sequence of vectors of dimension d vectors, and the function $a(\cdot)$ is the alignment function. The attention mechanism compares the degree of matching between the request vector and the object of attention, i.e..

$$e_i = a(u, v_i) \quad (3)$$

There are three commonly used alignment functions, dot product, product attention, and additive attention, calculated as shown in equation (4).

$$a(u, v_i) = \begin{cases} u^T v \\ u^T W_V \\ w_2^T \tanh(W_1 [u, v]) \end{cases} \quad (4)$$

After the attention mechanism is calculated, the normalized matching score vector is obtained. α_i , can judge the request vector u to the object V The calculation process is as follows:

$$\alpha_i = \frac{e_i}{\sum_i e_i} \quad (5)$$

Finally, the above can be summarised as a key-value mapping task, noting for inputs, respectively, as Q, K, V . The corresponding formula is as follows.

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V \quad (6)$$

Following the introduction of the attention mechanism, a detailed elucidation of the Transformer network is presented. The Transformer coding network is grounded in the attention mechanism, a feature that ensures equitable treatment of long-distance and short-distance dependencies in text data modeling. This approach effectively addresses the challenge of gradient disappearance, and the concurrent parallelization of computations contributes to a notable improvement in computational speed [28]. The specific structure is visually depicted in Figure 2:

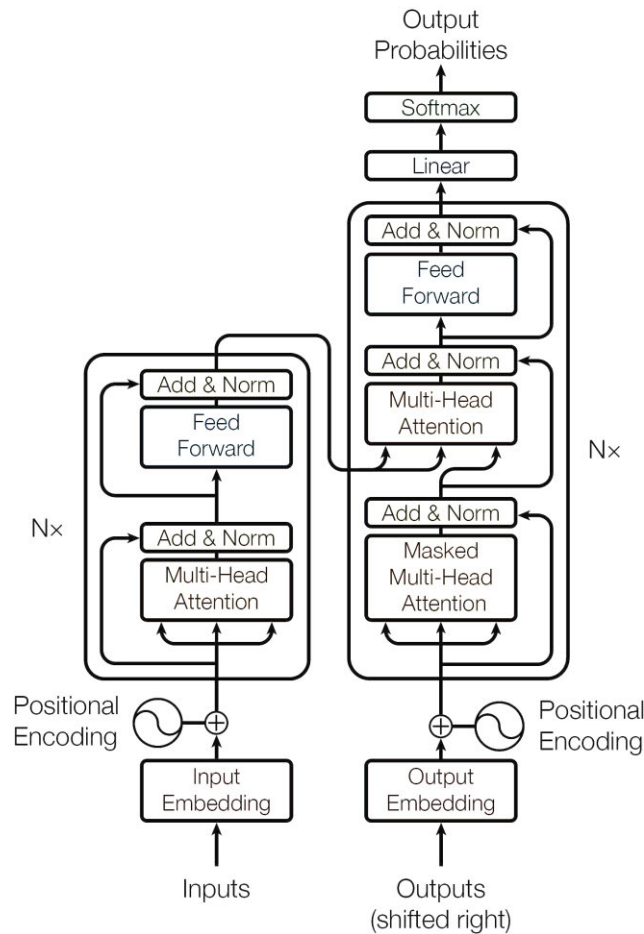


FIGURE 2. The framework for the transformer.

In Figure 2, it is evident that the Transformer’s encoder is assembled by stacking N identical substructures. These substructures encompass Multi-head Attention, Residual Connection, Layer Normalization, and Forward Network. Notably, the Multi-head Attention Network (MHAN) is founded on the fundamental attention network. It subdivides into distinct subspaces to glean diverse features from various perspectives. The computational process unfolds as follows:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h) W^O \quad (7)$$

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \quad (8)$$

where head_i is the learned attention representation in the i th subspace, and the final result is stitched together from the results of all the subspace operations.

C. MULTILEVEL FUSION PREDICTION BASED ON WORD VECTOR-CODING INTERACTION MA-TRANSFORMER

In this paper, an MA-Transformer model centered around text matching, is formulated by refining the CBOW word vector embedding method and the previously introduced Transformer model. The specific structure of this model is depicted in Figure 3:

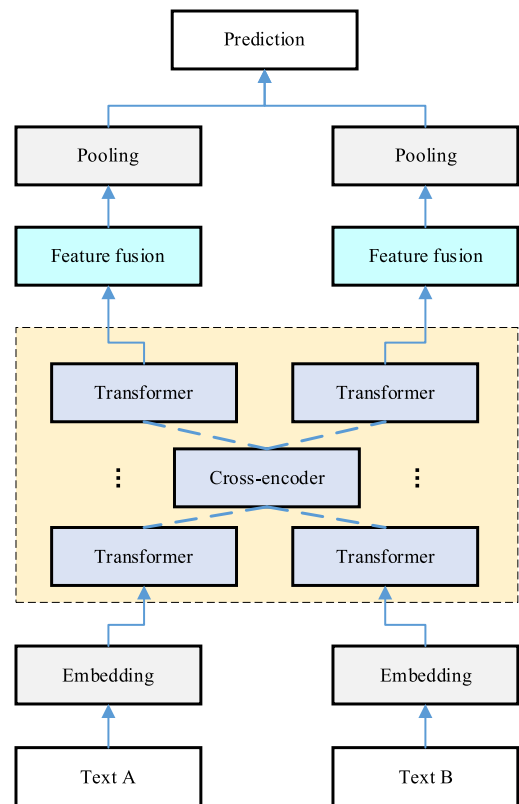


FIGURE 3. The framework for the proposed MA-Transformer.

In this model, the author has structured three layers, with the bottom layer comprising the text input and word vector embedding layer. Following this, there is an encoding layer situated between the two interconnected transformer layers, culminating in the topmost prediction layer. For the text matching task, the inputs consist of Text A and Text B, which can be represented as vectors, i.e., $A = \{A_1, A_2, \dots, A_n\}$ and $B = \{B_1, B_2, \dots, B_n\}$, in which the data in each set represents the meaning of its words. Following vector encoding in the embedding layer, the encoded vectors serve as input for the Transformer model. In the interactive connection encoding layer, the features are extracted from the two Transformer model columns, and within the encoder module, these features are fused. The fused features are then fed back into the Transformer module to accomplish feature interaction in the text matching process. Following the fusion of

interacting features, a self-encoding layer and a pooling layer are employed to integrate and pool the outputs from the feature extraction blocks, yielding the final vector representations for both texts.

The prediction layer leverages the final vector representations of the two texts as inputs to forecast the logical relationships between them. Post self-encoding and interaction coding layers, the model produces two output matrices: the semantic characteristics of Text A and the semantic interaction characteristics of Text B in relation to Text A. Both features are crucial information for the text-matching task. As previously analyzed, feature alignment is achieved through the application of residual concatenation during feature fusion, mitigating the risk of performance degradation due to excessive network depth.

After pooling the layers, we obtain text *A* and text *B*. With the final feature vectors of v_a and v_b , the goal of the prediction layer is to predict the logical relationship between text pairs based on the two feature vectors. A two-layer forward network and softmax function are used to classify and predict the enhanced features. It is shown as follows:

$$\hat{Y} = \text{softmax}(\text{gelu}(W_1^o v) W_2^o) \quad (9)$$

where v is the vector of both *A* and *B* after enhancement, including the difference between two separate columns of vectors and dot product change content. w_1 and w_2 are the weights of feature enhancement. The final output of the function is the prediction of the classification, i.e., to discriminate the correctness of the match between the two classes of text. In the classification prediction of the text question, the author uses the cross-entropy loss function as shown in equation (10):

$$\text{LossCrossEntropy} = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^C Y_{ij} \log \hat{Y}_{ij} \quad (10)$$

IV. EXPERIMENT RESULT AND ANALYSIS

A. EXPERIMENT SETUP

The primary objective of this paper is to undertake the text-matching task within the translation process. To accomplish this, classical paraphrase recognition datasets are chosen, wherein the same semantic scenario is expressed using either identical or different expressions. The selected datasets encompass two types: QQP [29] and AFQMC [30]. These datasets serve as the foundation for studying the textual content in the text matching process and investigating performance metrics within the retrieval-matching process, as shown in Table 2.

After completing the data selection and related tasks to determine whether we need to improve the performance of the model, according to the characteristics of the data and the deep learning classification research process of the data analysis mode, this paper selected precision, recall and F1-score as the model evaluation index. Firstly, precision measures the proportion of positive samples predicted by the model, which is suitable for situations with high false positive costs

TABLE 2. The information for the employed datasets.

Dataset	Information
QQP	Provided by Quora, which contains pairs of questions submitted by users on Quora, labelled with whether they are semantically similar or not. Binary labelling, where a label of 1 indicates that the two questions are semantically similar and a label of 0 indicates that they are not. Used for training and evaluating the performance of the model on the question matching task.
AFQMC	Provided by Ant Financial, a problem matching dataset for the financial domain. Binary labelling, with a label of 1 indicating that two questions are semantically similar, and a label of 0 indicating that they are not similar. Problem matching for the financial domain can be used to develop models to understand and deal with finance-related problems, e.g., in customer service to automatically answer questions posed by users.

and helps to improve the correlation of the results. Secondly, recall measures the proportion of actual positive samples correctly predicted by the model as positive and is suitable for situations with high false negative costs to ensure that the model does not miss any true positive examples. Finally, the F1 score is the harmonic mean of precision and recall, which takes into account both factors and is suitable for scenarios that require a balance between precision and recall. By comprehensively using these three indicators, the performance of the model can be evaluated more comprehensively, ensuring its effectiveness in different application scenarios. The specific calculation is shown in the formula: (11) - (13):

$$P = \frac{TP}{TP + FP} \quad (11)$$

$$R = \frac{TP}{TP + FN} \quad (12)$$

$$F1 = \frac{2 \times P \times R}{P + R} \quad (13)$$

where TP is the true positive, FP is the false positive, and FN represents the false negative.

To comprehensively assess the model's performance, the author opts for classic comparison methods commonly utilised in text-matching research. The selected methods for comparison include DSSM [31], DiSAN [32], ESIM [33], and BIMPM [24]. The specific introduction to the methods is concluded as follows: Developed by the Microsoft team, DSSM proposes a distributed representation of text to obtain feature vectors. It utilises these feature vectors to calculate the similarity between texts, particularly in information retrieval. The implementation of this framework is improved in this paper. Abandoning traditional RNN and CNN neural networks, DiSAN introduces a directional self-attention network text encoding method based on the attention mechanism. Extracting semantic information through the LSTM method, ESIM employs the attention mechanism to compute the similarity between texts and words. It is a classic method utilising LSTM as the interaction feature, making it a conventional approach to text matching. BIMPM completes semantic feature interaction from a multi-horizon perspective, leading to improved model performance. These comparison methods

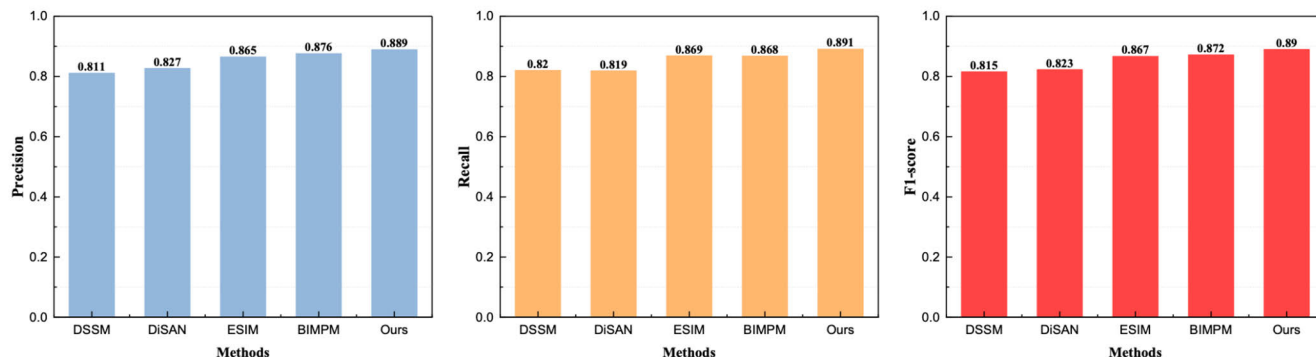


FIGURE 4. The comparison result on the QQP datasets.

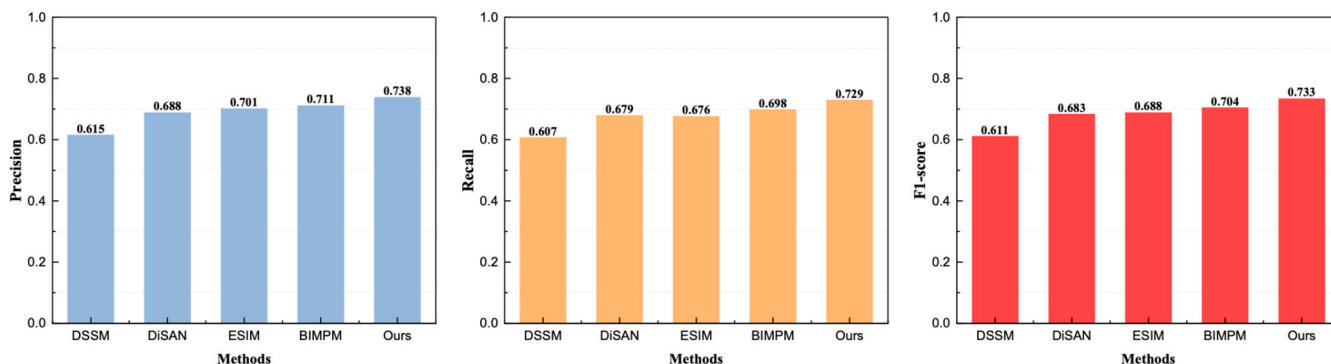


FIGURE 5. The comparison result on the AFQMC datasets.

are chosen to provide a comprehensive evaluation of the proposed model’s performance in the text-matching task. The experimental parameter settings for building the model in this article are as follows: the batch size of the dataset is 64, and the maximum sentence length in the batch is 50. The encoder has four layers; the hidden layer size is 256, the hidden layer size of the multi-head self-attention mechanism is 256, and the number of heads is 4. For the decoder, the embedding layer size is 256, and the number of layers is 4. The hidden layer size of the attention mechanism is 256, with four layers and four heads. This chapter uses the Adam optimizer for optimization methods. The initial value of the learning rate is $3e-4$, and the weight attenuation is $1e-5$. The learning rate attenuation adopts exponential attenuation, and the total number of training iterations is 100.

B. EXPERIMENT RESULT AND ANALYSIS

Following the selection of the experimental dataset and the definition of pertinent evaluation metrics and comparative methodologies, the author proceeded with the experiments. In this paper, the author adopted the algorithmic framework provided by PyTorch to construct the model. The experimental results were then calculated using two types of public datasets, as depicted in Figures 4 and 5, respectively.

Figure 4 illustrates the recognition results under the QQP dataset. Following the methodology introduced in Section IV-A, we tackled the challenge of matching text content labels for these datasets to derive the corresponding recognition results. It is evident that the proposed MA-Transformer method, as outlined in this paper, exhibits greater stability in precision and recall. The values for both indices stand at 0.889 and 0.891, respectively, highlighting the model’s balanced and elevated performance compared to current state-of-the-art text-matching algorithms.

The recognition results under the AFQMC dataset are depicted in Figure 5:

In Figure 5, it is observed that, under this dataset, the precision achieved by the proposed method is 0.738. While it falls short of reaching 0.8, considering the complexity of the Chinese text schema and matching task, this metric’s performance remains acceptable. Notably, it surpasses the metrics achieved by recent methods, which hover around 0.6. Furthermore, the F1 metric of the proposed method closely aligns with the precision and recall metrics, underscoring its significance for future analyses.

After completing model testing, the author proceeded to conduct ablation experiments on both dataset variants. These experiments involved reducing the embedding layer module, limiting the validation of the cross-module, excluding the

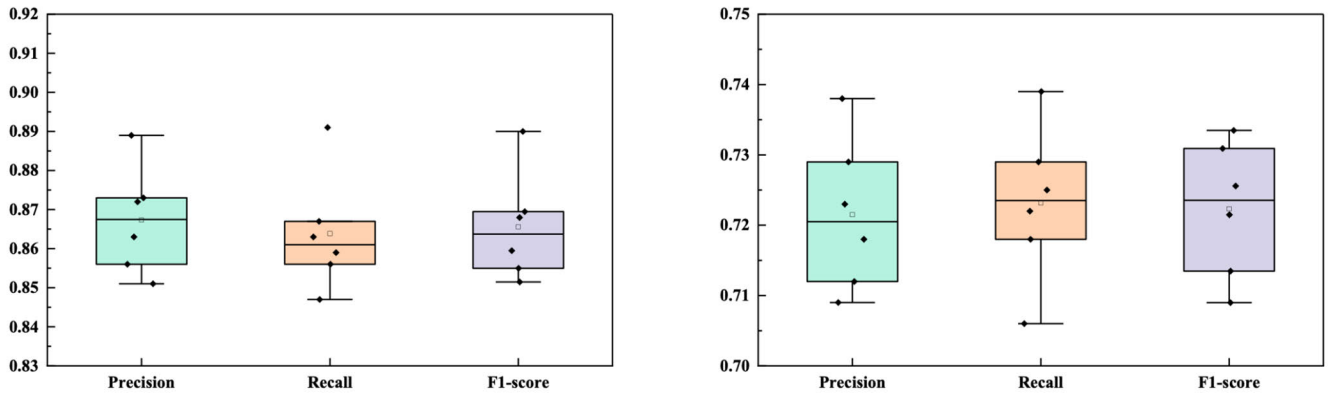


FIGURE 6. The ablation experiment on the QQP and AFQMC datasets.

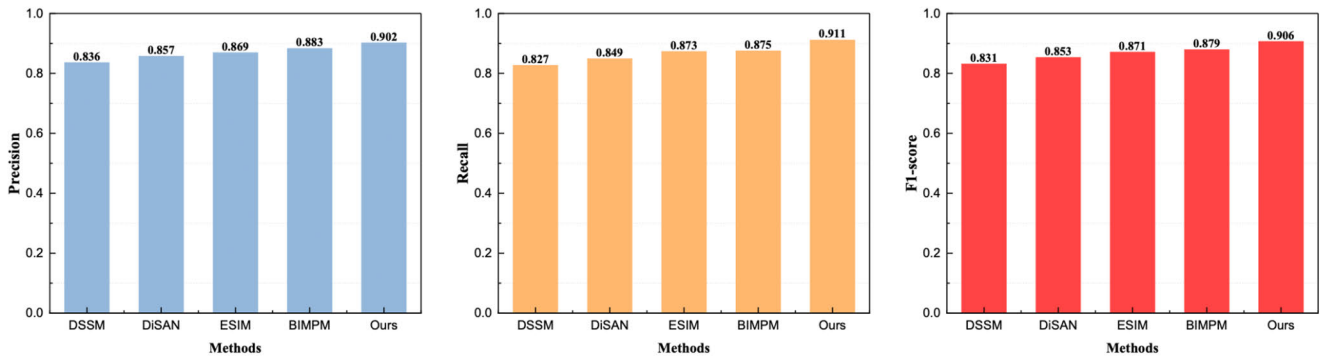


FIGURE 7. The comparison result on the established-translation dataset.

normalization layer, omitting the cross-encoder, and substituting the Transformer multi-module with LSTM. The results of these ablation experiments are presented in the following figures.

The results presented in Figure 6 highlight the robustness of the proposed method. The model’s construction through multi-layer semantics in its structure contributes to its consistent performance across various ablation experiments. The average precision achieved under the two types of datasets stands at 0.867 and 0.722, respectively. This underscores the rationality and effectiveness of the model construction.

C. THE PRACTICAL TEST FOR THE PROPOSED MODEL

Following the testing of the model on the public dataset, the author proceeded to evaluate the model using an actual dataset. This dataset was derived from real translation test exercises, encompassing student translations from the English elective course over the past five years, which were then compared to standard translations. The aim was to conduct a thorough analysis and comparison to achieve matching insights between the texts. The experimental test results are depicted in Figure 7:

In Figure 7, it is evident that the precision achieved by the proposed MA-Transformer model in this paper exceeds 90%.

This represents an impressive recognition result, especially considering the challenges posed by uncleaned sales sample data that often trouble other methods. The results affirm the advantages of the method proposed. Furthermore, the data results pertaining to recall and F1 scores underscore the superiority of the proposed method. Building upon this foundation, the author conducted further ablation experiments to analyse the model comprehensively, and the ablation process aligns with the comparisons made on public datasets in the previous section.

As depicted in Figure 8, the overall precision distribution is more dispersed due to the limited amount of data, ranging from the highest precision of 0.902 to the lowest of 0.881, which aligns with the classical BIMPM method. However, this dispersion is considered a side effect of the framework’s rationality. In contrast, the distribution of recall and F1 scores is more concentrated, yielding clearer and more consistent results. Therefore, the model proposed continues to yield satisfactory results in practical tests, showcasing its effectiveness even with limited data.

V. DISCUSSION

This paper addresses the text matching challenge within the realm of intelligent English translation, presenting the MA-Transformer, a text matching model based on multi-level

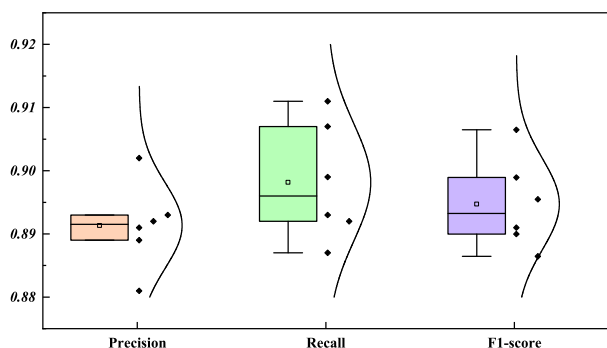


FIGURE 8. The ablation experiment on the established-translation dataset.

semantic feature extraction through the multi-level attention mechanism. The framework begins by enhancing the input features through the CBOV word embedding technique, extracting word vector features. Subsequently, the model undergoes further data analysis using an improved transformer model and related techniques, culminating in the completion of the matching task through feature enhancement and pooling. In the process of constructing the model, a semantic interaction feature extraction method based on the multi-head interaction attention mechanism is introduced. The interaction feature plays a pivotal role in the text matching model, enabling the mining of more detailed matching relationship features between pairs of texts based on semantics. The MA-Transformer method demonstrates commendable performance in comparison with methods like DSSM and ESIM.

Moreover, this paper leverages the transformer model for semantic feature extraction, showcasing significant advantages over CNN and RNN. The transformer model excels at handling long-distance dependencies through self-attention, enabling parallel computation to enhance training and inference speed. The introduction of positional encoding addresses sequence information concerns, ensuring good scalability. By integrating feature attention and word vector embedding, this paper achieves superior experimental results in the text-matching task.

Leveraging machine learning and deep learning technologies to address text matching issues in machine translation offers advantages such as contextual understanding, end-to-end learning, nonlinear modelling, large-scale data application, and transfer learning. These technologies provide powerful tools to enhance translation quality and adapt to diverse contexts. In practical applications, this model significantly improves translation accuracy and consistency through the robust processing capabilities of deep learning, better understanding complex contexts and semantic relationships, and adapting to different language styles and content.

In the future, deep learning is expected to bring higher translation quality, multimodal translation, and personalised translation, further enhancing the practicality and user experience of machine translation. However, it is essential to address challenges related to data privacy, social and cultural

differences, model transparency, and resource consumption to ensure the sustainable and responsible development of machine translation. Future research should focus on continuously improving deep learning models, strengthening multimodal translation research, emphasising transparency and interpretability, addressing cultural and social differences, prioritising data privacy and security, improving resource efficiency, and enhancing cross-disciplinary collaboration. These strategies will comprehensively advance machine translation technology, improve translation quality and sustainability, protect user privacy, respect cultural differences, and foster innovation and development in the field.

VI. CONCLUSION

This paper introduces an MA-Transformer text matching analysis framework based on multi-level semantic feature extraction to address the text matching challenges in English translation using artificial intelligence methods. The goal is to achieve higher accuracy in text-topic matching analysis for English teaching and learning. The framework integrates word vector embedding and multi-level transformer technology.

Tests conducted on two common paraphrase recognition matching task datasets, QQP and AFQMC, demonstrate that the framework outperforms single traditional recognition methods in terms of recognition accuracy and balanced performance. The recognition precision achieved is 0.889 and 0.738, respectively. In the actual model test, the precision of the framework reaches 0.902, surpassing commonly used models such as DSSM and ESIM. This offers methodological references and technical support for the intelligent matching of text in the design of future English translation systems.

Future research plans aim to expand the data processing capabilities of the current model to enhance text-matching abilities. This includes integrating text, speech, and other information to improve the model's overall performance through multimodal data fusion. Specifically, incorporating non-text information, such as speech and images, into the model can enhance its effectiveness in complex scenarios. Additionally, the introduction of advanced feature extraction techniques, such as deep semantic analysis based on pre-trained large-scale language models like GPT and BERT, will further improve the model's text understanding and matching capabilities. Establishing a standardised text-matching research dataset is also anticipated to provide references and technical support for a broader range of researchers, promoting overall advancement in the field. These improvements are expected to achieve higher matching accuracy and performance across more diverse application scenarios.

REFERENCES

- [1] S. R. Kudugunta, A. Bapna, I. Caswell, N. Arivazhagan, and O. Firat, "Investigating multilingual NMT representations at scale," 2019, *arXiv:1909.02197*.
- [2] K. Song, Y. Zhang, H. Yu, W. Luo, K. Wang, and M. Zhang, "Code-switching for enhancing NMT with pre-specified translation," 2019, *arXiv:1904.09107*.

- [3] Z. Chen, Y. Fu, Y. Zhang, Y.-G. Jiang, X. Xue, and L. Sigal, "Multi-level semantic feature augmentation for one-shot learning," *IEEE Trans. Image Process.*, vol. 28, no. 9, pp. 4594–4605, Sep. 2019.
- [4] B. Chen, M. Xia, and J. Huang, "MFANet: A multi-level feature aggregation network for semantic segmentation of land cover," *Remote Sens.*, vol. 13, no. 4, p. 731, Feb. 2021.
- [5] F. A. Acheampong, H. Nunoo-Mensah, and W. Chen, "Transformer models for text-based emotion detection: A review of BERT-based approaches," *Artif. Intell. Rev.*, vol. 54, no. 8, pp. 5789–5829, Dec. 2021.
- [6] X. Shi, Z. Wang, H. Zhao, S. Qiu, R. Liu, F. Lin, and K. Tang, "Threshold-free phase segmentation and zero velocity detection for gait analysis using foot-mounted inertial sensors," *IEEE Trans. Human-Mach. Syst.*, vol. 53, no. 1, pp. 176–186, Feb. 2023.
- [7] Y. Bengio, H. Schwenk, and J. S. Senécal, "Neural probabilistic language models," *Stud. Fuzziness Soft Comput.*, vol. 194, pp. 137–186, Jan. 2006.
- [8] M. L. Forcada and L. Rpiñeco, *Recursive Hetero-Associative Memories for Translation*. Berlin, Germany: Springer, 1997.
- [9] M. T. Luong, I. Sutskever, Q. V. Le, O. Vinyals, and W. Zaremba, "Addressing the rare word problem in neural machine translation," 2014, *arXiv:1410.8206*.
- [10] M. Junczys-Dowmunt, T. Dwojak, and H. Hoang, "Is neural machine translation ready for deployment? A case study on 30 translation directions," 2016, *arXiv:1610.01108*.
- [11] Y. Wu, M. Schuster, and Z. Chen, "Google's neural machine translation system: Bridging the gap between human and machine translation," 2016, *arXiv:1609.08144*.
- [12] J. Gehring, M. Auli, and D. Grangier, "Convolutional sequence to sequence learning," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 1243–1252.
- [13] R. Sennrich, B. Haddow, and A. Birch, "Improving neural machine translation models with monolingual data," 2015, *arXiv:1511.06709*.
- [14] S. Ren, W. Chen, and S. Liu, "Triangular architecture for rare language translation," 2018, *arXiv:1805.04813*.
- [15] A. Aizawa, "An information-theoretic perspective of tf-idf measures," *Inf. Process. Manag.*, vol. 39, no. 1, pp. 45–65, 2003.
- [16] S. Robertson, H. Zaragoza, and M. Taylor, "Simple BM25 extension to multiple weighted fields," in *Proc. 13th ACM Int. Conf. Inf. Knowl. Manage.*, Nov. 2004, pp. 42–49.
- [17] S. T. Dumais, "Latent semantic analysis," *Annu. Rev. Inform. Sci. Technol.*, vol. 38, no. 1, pp. 189–230, 2004.
- [18] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Jan. 2003.
- [19] P.-S. Huang, X. He, J. Gao, L. Deng, A. Acero, and L. Heck, "Learning deep structured semantic models for web search using clickthrough data," in *Proc. 22nd ACM Int. Conf. Conf. Inf. Knowl. Manage. CIKM*, 2013, pp. 2333–2338.
- [20] R. Paul, J. Arkin, N. Roy, and T. M. Howard, "Efficient grounding of abstract spatial concepts for natural language interaction with robot manipulators," in *Proc. Robot., Sci. Syst. XII*, 2016.
- [21] H. Palangi, *Deep Learning for Sequence Modelling: Applications in Natural Languages and Distributed Compressive Sensing*. Vancouver, BC, Canada: Univ. British Columbia, 2017.
- [22] A. P. Parikh, O. Täckström, D. Das, and J. Uszkoreit, "A decomposable attention model for natural language inference," 2016, *arXiv:1606.01933*.
- [23] Q. Chen, X. Zhu, Z. Ling, S. Wei, H. Jiang, and D. Inkpen, "Enhanced LSTM for natural language inference," 2016, *arXiv:1609.06038*.
- [24] Z. Wang, W. Hamza, and R. Florian, "Bilateral multi-perspective matching for natural language sentences," 2017, *arXiv:1702.03814*.
- [25] T. Kenter, A. Borisov, and M. de Rijke, "Siamese CBOW: Optimizing word embeddings for sentence representations," 2016, *arXiv:1606.04640*.
- [26] Y. Feng, C. Hu, H. Kamigaito, H. Takamura, and M. Okumura, "A simple and effective usage of word clusters for CBOW model," *J. Natural Lang. Process.*, vol. 29, no. 3, pp. 785–806, 2022.
- [27] Z. Liu, Q. Lv, Z. Yang, Y. Li, C. H. Lee, and L. Shen, "Recent progress in transformer-based medical image analysis," *Comput. Biol. Med.*, vol. 164, Sep. 2023, Art. no. 107268.
- [28] Y. Liu, Y. Zhang, Y. Wang, F. Hou, J. Yuan, J. Tian, Y. Zhang, Z. Shi, J. Fan, and Z. He, "A survey of visual transformers," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 6, pp. 7478–7498, Jun. 2024.
- [29] K. H. Alyoubi, F. S. Alotaibi, A. Kumar, V. Gupta, and A. Sharma, "A novel multi-layer feature fusion-based BERT-CNN for sentence representation learning and classification," *Robotic Intell. Autom.*, vol. 43, no. 6, pp. 704–715, Nov. 2023.
- [30] M. S. Chen, J. Lee, and H. Z. Ye, "Data-efficient machine learning potentials from transfer learning of periodic correlated electronic structure methods: Liquid water at AFQMC, CCSD, and CCSD (T) accuracy," *J. Chem. Theory Comput.*, vol. 19, no. 14, pp. 4510–4519, 2023.
- [31] Y. Shen, X. He, J. Gao, L. Deng, and G. Mesnil, "Learning semantic representations using convolutional neural networks for web search," in *Proc. 23rd Int. Conf. World Wide Web*, Apr. 2014, pp. 373–374.
- [32] T. Shen, T. Zhou, and G. Long, "DiSAN: Directional self-attention network for RNN/CNN-free language understanding," in *Proc. AAAI Conf. Artif. Intell.*, vol. 32, no. 1, 2018, pp. 5446–5455.
- [33] H. Rebecq, D. Gehrig, and D. Scaramuzza, "ESIM: An open event camera simulator," in *Proc. Conf. Robot Learn.*, 2018, pp. 969–982.



HUI YANG received the B.A. degree in English from Anhui Agricultural University, Anhui, in 2006, and the Master of Arts degree in English language and literature from Nankai University, Tianjin, in 2008. Since 2008, she has been a Teacher with the School of Foreign Languages, Anhui Agricultural University. Her research interests include translation theory and practice, intercultural communication, and artificial intelligence translation. She has been awarded the title of "Rising Star in Provincial Teaching Circles of Anhui Province."

• • •