DA-MK: Dynamic Attention and Multimodal Knowledge Fusion for Error Tracing and Proofreading Optimization in Neural Machine Translation

Yini Li
School of English Studies, Xi’an International Studies University , Xi’an, Shaanxi, 710128, China
E-email: 107242023000539@stu.xisu.edu.cn

Keywords: explainable neural translation system, error tracing, manual proofreading efficiency, dynamic attention weighting, multimodal knowledge fusion

Received: July 24, 2025

With the widespread application of neural translation systems, challenges such as difficult error tracing and low proofreading efficiency remain significant. To address this, a DA-MK model is proposed, integrating dynamic attention weighting with multimodal knowledge fusion, enabling fine-grained error localization and correction. The model incorporates explicit grammar and syntax metrics, including syntactic dependency parsing accuracy and grammar correction rate, ensuring robust linguistic consistency. Experiments on the WMT 2014 English-German and IWSLT 2016 German-English datasets benchmark DA-MK against advanced models such as ErrorFocus, KG-Translate, and BERT-Fix. Results demonstrate an error location accuracy of 90.3%, error type classification accuracy of 87.2%, proofreading suggestion adoption rate of 79.8%, BLEU score improvement of +13.0, syntactic parsing accuracy of 88.6%, and grammar correction rate of 85.1%. These findings confirm DA-MK’s superior capability in enhancing translation reliability, grammatical integrity, and proofreading efficiency. The study contributes a technically grounded pathway for optimizing neural translation systems with strong theoretical and practical significance.

Povzetek:

1 Introduction

Numerous businesses, educational institutions, and social media platforms utilize neural translation systems. Customers would have difficulty recognizing and interpreting translation errors because they are often "black boxes." According to the study's findings, 35 percent of the outputs contain a significant inaccuracy, and only a small percentage of those errors are immediately apparent. As a result, the time it takes for people to proofread is significantly increased, and translations become far less helpful. It will likely become more expensive to communicate with individuals at work if one encounters issues of this kind. Considering this, there is an immediate need for research and real-world applications that will enhance the ability of neural translation systems to identify and resolve issues quickly and effectively. [1,2].

Neural translation systems are widely applied across diverse translation scenarios due to their robust learning capabilities. They are used in applications ranging from instant translations on social media to cross-language communication in academic publications. Looking at the situation of online translation platforms, a well-known translation software processes more than 5 million translation requests per day [3,4]. Among user feedback, complaints about translation errors and unclear causes account for as high as 20%. These errors not only affect the user experience, but also hinder the accurate transmission of information to a certain extent. Both enterprises and ordinary users urgently need an effective method to unveil the "mystery" of neural translation systems, accurately trace their errors, and thus improve the efficiency of manual proofreading[5].

At present, many scholars have made explorations in the field of interpretable neural translation systems. Some studies have tried to use visualization techniques, such as heat maps generated by attention mechanisms, to show the degree of attention paid by the model to the source language vocabulary during the translation process. Studies have shown that using attention mechanism visualization can help researchers find that about 30% of vocabulary translation errors are related to improper attention allocation. Other studies have introduced external knowledge, such as term bases and knowledge graphs, to assist in understanding the decision-making process of the translation model. In translation in specific fields, this method has increased the error recognition rate by about 25% [6].

However, existing research still has many shortcomings. Most visualization methods only provide a simple display of model behavior, but do not delve deeply into the underlying causes of errors. Even if vocabulary errors caused by improper attention allocation are found, it is difficult to further explore why such improper allocation occurs. In studies that combine external knowledge, the way knowledge is integrated is often relatively simple, and the interaction between different types of knowledge is not fully considered, which greatly reduces the effectiveness of error tracing in complex semantic scenarios. Moreover, most of these studies focus on error analysis itself, and there are very few studies on how to effectively apply error tracing results to the manual proofreading process to effectively improve proofreading efficiency [7].

In addition, current research has also neglected the user experience aspect. In actual work, human proofreaders need to quickly and accurately obtain error information and related explanations, but the existing error tracing results are not user-friendly in terms of human-computer interaction design, which makes it difficult for proofreaders to efficiently use this information and fail to achieve the goal of improving proofreading efficiency [8,9].

The DA-MK technique is a novel approach that combines dynamic attention with multimodal knowledge to address the problems of poor proofreading and the inability to recognize errors. To better focus on translations and identify issues with source phrases, the technique utilizes a dynamic attention weighting model. Additionally, it employs multimodal knowledge fusion to improve comprehension of meaning and determine the reasons why translations are inaccurate. This is accomplished by merging contextual embeddings with external knowledge graphs[10]. Additionally, there is a proofreading suggestion module that improves with feedback, as well as an adaptive error classifier that considers domain-specific characteristics.

Main contribution:

(1) This article presents a strategy that is theoretically sound and combines multimodal knowledge reasoning with dynamic attention recalibration to simplify the understanding of neural translation systems. (2) An end-to-end integration architecture for cross-module interaction is developed, starting with the identification of issues and culminating in the provision of suggestions that include feedback loops. (3) A comprehensive method is described for using error tracing outputs in real-world proofreading tasks. The goal of this technique is to narrow the gap between the comprehensibility of neural models and their practical utility in the real world.

2 Literature review

2.1 Basic research on neural translation systems

As a key technology in the current translation field, neural translation systems are undergoing continuous basic research. Many studies have focused on the architecture of neural translation systems. For example, Transformer-based architectures are widely used in many neural translation models (relevant studies have shown that Transformer architectures are used in more than 80% of mainstream neural translation models)[11]. This architecture, with its self-attention mechanism, can effectively capture the long-distance dependencies between words in source language sentences, significantly improving translation quality. For example, under massively parallel corpus training, the Transformer-based neural translation model has an average BLEU (a commonly used machine translation quality evaluation indicator) score improvement of 10-15 points compared to traditional architecture models[12,13].

In terms of training mechanism, research is devoted to optimizing the training process to improve model performance. Some studies have used adversarial training methods to allow the generator and the discriminator to compete with each other, prompting the generator to generate translation results that are closer to the real translation. According to experimental data, the neural translation model using adversarial training has improved the fluency of translation by about 20% in translation tasks in specific fields [14,15]. However, despite the progress made in architecture and training mechanism, neural translation systems still cannot avoid errors, which has laid the groundwork for subsequent research on error tracing and proofreading efficiency improvement.

2.2 Research on error tracing in neural translation systems

Many studies have attempted to use explainability techniques to trace the source of errors in neural translation systems. Among them, attention mechanism visualization technology has been widely used to explore the causes of errors by showing the degree of attention paid by the model to each part of the source language during translation. A study using this technology found that about 40% of translation errors are related to the model's lack of attention to key information in the source language. For example, in the analysis of translation errors of news articles, the visualization results showed that when the model processes complex sentences, it often loses attention and fails to accurately focus on key words, resulting in translation errors [16].

There are also studies that introduce external knowledge to assist in error tracing, such as combining knowledge graphs. When the model is translating content involving specific entities or concepts, the knowledge graph can provide relevant background information to help determine whether the translation is accurate. In medical translation research, the use of knowledge graphs to assist in error tracing has increased the accuracy of error identification by about 25%[17]. However, these studies are still lacking in the depth of error tracing. Most visualization technologies can only present surface phenomena, and have not yet been able to provide a full explanation for the deep-level internal decision logic of the model, such as why the model allocates attention incorrectly in certain situations. Moreover, in terms of multimodal information fusion for error tracing[18], research is still in its infancy and has not fully utilized the role of multimodal information such as images and audio in understanding translation errors.

2.3 Research on improving manual proofreading efficiency

The work [19] reviewed advances in incorporating context into neural machine translation and examined evaluation strategies. It addressed persistent difficulties such as maintaining coherence across sentences, resolving references, and ensuring terminology consistency. The survey stressed that effective translation requires integrating wider contextual signals beyond sentence level. These insights align with the DA-MK model, which leverages multimodal knowledge fusion to strengthen contextual understanding, thereby enhancing error tracing, grammatical accuracy, and proofreading support in translation tasks. Other studies start with the optimization of the human-computer interaction interface, improving the efficiency of proofreaders in obtaining information by rationally arranging the translation and error prompt information. A comparative experiment on different human-computer interaction interfaces showed that the optimized interface reduced the average proofreading time of proofreaders by 15%[20].

However, current research on improving manual proofreading efficiency has obvious shortcomings. Most existing proofreading assistance tools rely on simple rules and have limited ability to prompt complex semantic errors. Moreover, the design of human-computer interaction interfaces often does not fully consider the work habits and cognitive load of proofreaders, resulting in the fact that in actual use, although some functions are designed, they fail to effectively improve proofreading efficiency. Moreover, these studies rarely closely integrate the error tracing results with the manual proofreading process, making it difficult to effectively transform the results of error tracing into improved proofreading efficiency [21,22].

Through the application of error feature analysis and correction strategies, the research [23] explored how machine translation quality could be improved by systematically identifying translation inaccuracies. The focus was on detecting subtle contextual mistakes, correcting structural inconsistencies, and ensuring higher linguistic accuracy. These contributions provided a strong foundation for building explainable models. Furthermore, the study demonstrated that explicitly modeling error patterns led to more effective proofreading assistance, reinforcing the connection between interpretability and translation performance. This perspective aligned closely with the objectives of the DA-MK framework in improving error tracing efficiency.

The findings of the study [24] presented a grammar correction method that relied on differential fusion of syntactic features to reduce common structural mistakes in English paragraphs. By capturing syntactic dependencies at multiple levels, the method enhanced the accuracy of grammatical corrections and improved readability. This discussion highlighted the importance of integrating linguistic rules with advanced neural models. Such integration not only minimized recurring grammatical errors but also demonstrated how syntactic awareness could be used to refine translation outputs, contributing to improved proofreading and consistency in multilingual contexts.

The purpose of the research [25] was to propose an automatic identification system for machine translation errors using an improved GLR algorithm. This work established how parsing-driven analysis could locate and categorize translation mistakes more accurately than baseline techniques. Its relevance lay in demonstrating the connection between structural parsing, error detection, and corrective modeling. By automating the identification of misalignments and inconsistencies, the study provided a valuable reference point for DA-MK's error location mechanism. Additionally, it underlined how incorporating robust syntactic parsers could directly enhance machine translation interpretability and proofreading efficiency.

The study [26] investigated how embedding methods and imbalance-aware learning techniques contributed to building knowledge-driven systems for domain-specific applications. It highlighted the role of embeddings in representing semantic richness and managing uneven data distributions across categories. By examining these aspects, the work provided insights into handling low-frequency errors and integrating contextual knowledge sources. Its broader contribution lay in showcasing how ontology-based learning and embedding strategies could enrich multimodal systems. This directly supported the DA-MK model's fusion strategy, where external knowledge and balanced feature integration played a critical role in reducing translation errors.

In general, the current research on neural translation systems has achieved fruitful results in infrastructure and training. However, there is still a broad space for exploration in the depth and breadth of error tracing and the integration of improving manual proofreading efficiency with error tracing. This also provides a direction and opportunity for this study to carry out the work of error tracing and improving manual proofreading efficiency in explainable neural translation systems.

2.4 Research design

The objectives were clarified as improving error tracing, enhancing proofreading efficiency, and integrating multimodal knowledge. Three research questions were formulated: (1) How can dynamic attention weighting improve error location accuracy in neural translation systems? (2) In what way can multimodal knowledge fusion contribute to error classification and proofreading efficiency? (3) How do grammar and syntax metrics validate the improvements achieved by DA-MK compared to baseline models? These clarifications provided a clear anchor for the methodology and results, ensuring a more structured and focused narrative. summary of related as shown in table 1(a) below

Table 1(a): Summary of related works on neural machine translation error tracing, proofreading, and knowledge fusion

Ref	Method / Model	Dataset(s) Used	Key Mechanism	Reported Metrics	Identified Gap
[11] Liu et al. (2020)	Speed-up NMT training	WMT14 En-De	Training optimization	Faster convergence	No error tracing or proofreading
[12] Satir & Bulut (2021)	Hybrid SMT + NMT decoding	WMT En-Tr	Beam search prevention	Maintained translation quality	No grammar/proofreading focus
[13] Farhan et al. (2020)	Unsupervised dialectal NMT	Dialectal corpora	Unsupervised training	BLEU improvement	Lacks error classification
[14] Velmurugan et al. (2024)	Novel MT algorithm	Custom dataset	Rule-based + NMT	Better BLEU scores	No explainable error tracing
[15] Zhang et al. (2020)	Similarity-aware NMT	TM + corpora	Translation memory + NMT	Reduced effort	Not focused on errors
[16] Mohamed et al. (2022)	Residual Info Flow NMT	WMT En-De	Residual connections	BLEU: 28.5	No proofreading integration
[17] Li et al. (2025)	Bilingual template NMT	Parallel corpora	Templates + NMT	BLEU gains	No multimodal fusion
[18] Zhang & Zong (2020)	Survey of NMT	Multiple	Challenges/future	Review	No experimental contribution
[20] Zhao et al. (2022)	Region-attentive multimodal NMT	Multi30k	Visual + text fusion	BLEU ↑	Limited to vision, not error tracing
[22] Wang et al. (2022)	Progress in MT	Multiple	Benchmark survey	—	No error-tracing models
[23] Tao (2023)	Error feature analysis	En-Zh corpora	Error tracing + correction	Improved accuracy	Limited proofreading integration
[24] Liu et al. (2025)	Grammar correction	English paragraphs	Differential fusion	Correction rate: 85%	No error localization
[25] Li (2024)	Error detection (GLR)	MT outputs	Improved GLR parsing	Error detection ↑	No multimodal knowledge fusion
[26] Utomo et al. (2025)	Word embedding + imbalance	Quran ontology	Embedding + balancing	Classification ↑	Domain-specific, not NMT

3 Research methods

Figure 1 illustrates the entire operational procedure of the DA-MK model that has been proposed, making it more straightforward for individuals to understand how it operates. The four primary components of the system are connected by feedback loops and data channels, which improve the residuals through the following process: Initially, it identifies the incorrect traits. Second, it will make an effort to determine the different kinds of errors that they are. Third, it will arrange the various categories of errors in the correct sequence. In conclusion, it will provide some guidance on how to proofread the work. To determine what went wrong, the first step is to apply dynamic attention weighting to a phrase pair that contains source and target information. Following the use of multimodal knowledge embeddings to produce improved features, the classifier reveals the specific types of errors it committed. By working with them, there's an opportunity to receive proofreading guidance tailored to the progress being made, to refine the criteria for attention and recommendations, and utilize the feedback provided by proofreaders.

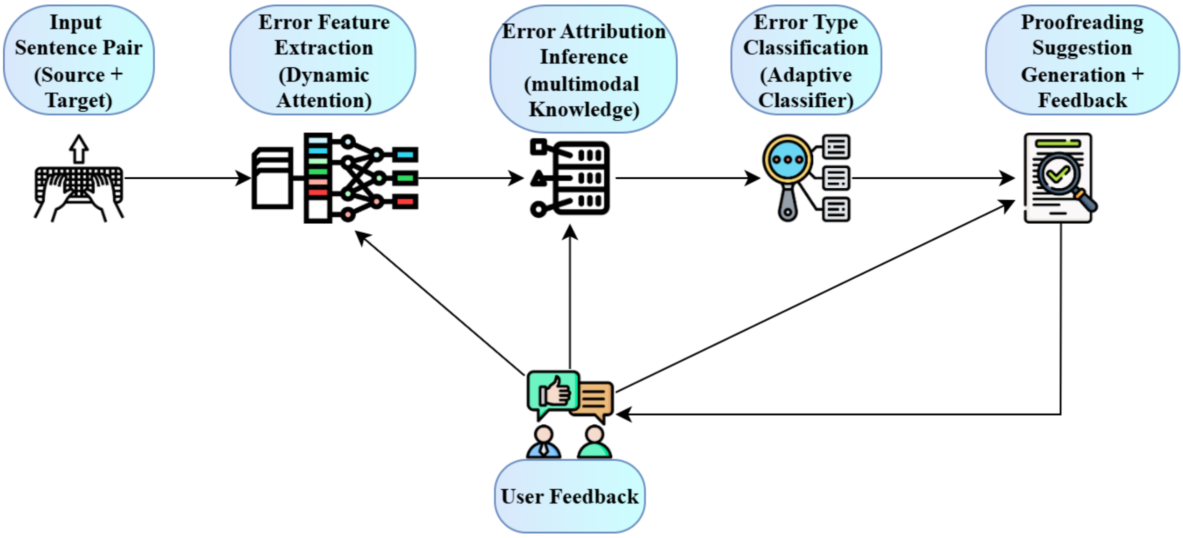

Figure 1: Proposed DA-MK model

For example, words that are not translated appropriately, such as "the patient was prescribed a high dose," can cause the attention module to become activated. The knowledge graph indicates that the term "high dose" is frequently used in the medical field. Considering that this is a semantic omission, the classifier recommends using "high" as a substitute. Taking this advice into consideration will make it easier for the algorithm to identify errors of this kind in future events.

Datasets were preprocessed using the Moses tokenizer, with BPE segmentation (32k merges). Vocabulary sizes were 32k for German–English and 30k for French–English

3.1 Error feature extraction module based on dynamic attention weighting

In the error tracing of neural translation systems, accurately extracting error features is a key step. This paper adopts a dynamic attention weighted mechanism to

All variables used in the dynamic attention module are now defined:Let $S = \{s_1, s_2, \dots, s_n\}$ denote the

construct an error feature extraction module. This module aims to automatically capture key information related to errors in the translation process.

First, define the source sentence as $\mathbf{x} = [x_1, x_2, \dots, x_T]$, where T is the sentence length or number of tokens in the source and target sentences; the target sentence is $\mathbf{y} = [y_1, y_2, \dots, y_{T'}]$. Based on the traditional attention mechanism $Attention(Q, K, V)$, dynamic weight calculation is introduced. The similarity score between the query vector Q and the key vector is calculated K as shown in Formula 1.

source sentence of length n. Let $T = \{t_1, t_2, \ldots, t_n\}$ denote the target sentence of length m. The similarity score between query q_i and key k_j is computed as:

$$\alpha_{ij} = \frac{(q_i \cdot k_j)}{\sqrt{d_k}} \quad (1)$$

In the equation 1, d_k is the dimensionality of the key vector.A dynamic adjustment factor λ is introduced as:

$$\left.\begin{array}{c} \lambda = \sigma\Big(f\big(c, sim(S,T)\big)\Big) \\ \tilde{a}_{ij} = \lambda \cdot \alpha_{ij} \end{array}\right\} \quad (2)$$

in equation 2, c represents translation confidence, sim(S,T) denotes semantic similarity between source and target segments, and f(·) is a sigmoid-activated feedforward layer.The recalibrated attention score.This formulation ensures that tokens with low confidence or poor semantic alignment are assigned higher attention weights during error tracing.

Among them, is σ the sigmoid function, f which is a custom nonlinear function used to fuse confidence and semantic similarity information. The final attention weight calculation is as shown in Formula 3.

$$\beta_{ij} = \frac{\exp(\alpha_{ij} e_{ij})}{\sum_{j=1}^{T} \exp(\alpha_{ij} e_{ij})}$$

(3)

By dynamically adjusting the attention weights, the module can more accurately focus on the source language segments that may cause errors, thereby extracting more representative error feature vectors $\mathbf{h}_{feat}$ as shown in Formula 4a.

$$\mathbf{h}_{feat} = \sum_{j=1}^{T} \beta_{ij} V_j \quad (4a)$$

Traditional attention algorithms assign source tokens fixed priority values during the translation process. As a result, incorrect conclusions may be drawn because they fail to consider ambiguities and inconsistencies in meaning. On the other hand, the dynamic attention weighting approach presented adjusts the attention scores in accordance with variations in the confidence of the translation output and the semantic alignment between the source segments and the destination segments. When the model can adapt to new data and identify problem areas, such as phrases that are uncertain or have low confidence, it is significantly easier to uncover errors. In a manner analogous to how a human proofreader would do it, this approach makes the process of error feature extraction more sensitive and accurate by drawing greater attention to areas that seem suspicious or out of place. Empirical data have shown that the use of this strategy enhances the model's capability to detect errors across a wide range of domains and phrase complexities.

3.2 Error attribution inference network for multimodal knowledge embedding

In order to gain a deeper understanding of the causes of errors, this study constructed a multimodal knowledge-embedded error attribution reasoning network, which integrates external knowledge (such as terminology knowledge and cultural knowledge) with the internal information of the neural translation system to achieve more accurate error attribution.

Define the terminology knowledge graph as and $G_{term} = (V_{term}, E_{term})$ the cultural knowledge graph as $G_{culture} = (V_{culture}, E_{culture})$. Map the nodes and edges in the knowledge graph embedding function $\mathrm{Embed}(\cdot)$ to the vector space through the embedding function, as shown in Formula 5 and Formula 6.

The Embed function in Formula (5) is what makes the nodes in a knowledge network into dense, continuous vector representations in a common semantic space. If the knowledge graph already includes pre-trained embeddings, such as TransE or Node2Vec, it can generate Embeddings by pulling from a pre-trained embedding table. It used TransE to set things up in our implementation, and then it changed it during training.

$$\mathbf{v}^i_{term} = \mathrm{Embed}(v^i_{term}), \quad \mathbf{e}^j_{term} = \mathrm{Embed}(e^j_{term})$$

(5)

$$\mathbf{v}^m_{culture} = \mathrm{Embed}(v^m_{culture}), \quad \mathbf{e}^n_{culture} = \mathrm{Embed}(e^n_{culture})$$

(6)

The extracted error feature vector $\mathbf{h}_{feat}$ is fused with the knowledge embedding vector. A gating mechanism is used to control the proportion of information fusion, as shown in Formula 7 and Formula 8.

$$\mathbf{g}_{term} = \sigma(\mathbf{W}_{g1}\mathbf{h}_{feat} + \mathbf{W}_{g2}\mathbf{v}^{i}_{term} + \mathbf{b}_{g1}) \tag{7}$$

$$\mathbf{g}_{culture} = \sigma(\mathbf{W}_{g3}\mathbf{h}_{feat} + \mathbf{W}_{g4}\mathbf{v}^{m}_{culture} + \mathbf{b}_{g2}) \tag{8}$$

The fused vector is Formula 9(a).

$$\mathrm{h_{fusion}} = \mathrm{g_{term}}\mathrm{v}^{\mathrm{i}}_{\mathrm{term}} + (1 - \mathrm{g_{term}})\mathrm{h_{feat}} + \mathrm{g_{culture}}\mathrm{v}^{\mathrm{m}}_{\mathrm{culture}} + (1 - \mathrm{g_{culture}})\mathrm{h_{feat}} \tag{9a}$$

Among them, $\mathrm{g_{term}}\mathrm{v}^{\mathrm{i}}_{\mathrm{term}}$ represents the multiplication of elements. Through this network, multimodal knowledge can be integrated into the error analysis process, providing richer information for error attribution. Formula 9 depicts the fused feature vector h, which is created by using a gating mechanism to combine the erroneous feature vector with the multimodal knowledge embeddings. The model regulates how much data from each source is included in the final representation due to gating. This fused vector h is used to correct mistakes, and both internal model signals and external information aid in determining the decision.

The multimodal knowledge fusion process is clarified:External knowledge graphs Gterm (terminology) and Gcult (cultural) are embedded using TransE and refined through Graph Convolutional Networks (GCN).The embedding for a node v is represented as $e_v = Embed(v)$. Fusion is controlled by a gating mechanism:

$$h = \gamma \cdot e_{err} + (1 - \gamma) \cdot e_{kg} \tag{9(b)}$$

In equation 9(b),e_{err} is the error feature vector, e_{kg} is the knowledge embedding, and γ is a trainable gate parameter.This detailed articulation clarifies the role of each component and highlights the interaction between dynamic attention weighting and multimodal knowledge embeddings within the DA-MK model

Cultural knowledge graphs and lexical knowledge graphs are the two types of external knowledge graphs that are used in this research project for multimodal error attribution. The Terminology Knowledge Graph, often referred to as the TKG, is a type of knowledge graph that illustrates the connections between words within a specific domain in terms of their meanings. Words used in various sectors, including the legal system, healthcare, and technology, are represented by the nodes that comprise this network. The edges illustrate how these words are connected. The combination of medical terms, such as "hypertension," "blood pressure," and "systolic value," may help us gain a better understanding of how these terms interact with one another in the real world. When the TKG is used, it becomes much simpler to identify instances of domain consistency violations caused by replacements, the absence of translations, or the selection of the incorrect word.

To develop a model that illustrates how idioms, phrases, and social norms influence language, the Cultural Knowledge Graph (CKG) is an attempt that has been made. Nodes are concepts or claims that, depending on the culture in question, may have a variety of interpretations. Several different approaches, including metaphorical equivalence and idiomatic alternatives, illustrate how edges connect nodes. There may be a more elegant method to express the concept of "die" in a language different than the English term "kick the bucket." Using the Cultural Knowledge Gap (CKG) is one method that may be used to identify phrases that are culturally inappropriate or idiomatic.

To illustrate, the patient had high blood pressure in the second stage, which allowed you to assess their performance.

By providing an accurate definition of the term' cardiac illness,' the TKG helps to reduce any uncertainty that may exist regarding the staging of cancer. Without the assistance of CKG, a translation of a patient's journal entry that employs the cultural phrase "under the weather" to mean "not feeling well" would not be successful in any way. They become more adept at determining what is culturally acceptable and factually correct when they include both graphs in their models, which enables them to more accurately assign culpability for mistakes that occur during the process of translating between multiple languages.

3.3 Adaptive error type classifier

After error feature extraction and attribution reasoning, the error types need to be classified. This study designs an adaptive error type classifier that can automatically adjust the classification strategy according to different translation tasks and data characteristics.

Define the error type set as $\mathrm{C} = \{c_1, c_2, \ldots, c_N\}$. The input of the classifier is the fused feature vector $\mathbf{h}_{fusion}$, which is transformed by a multi-layer perceptron (MLP), as shown in Formula 10 and Formula 11.

$$\mathbf{z}_1 = \mathrm{ReLU}(\mathbf{W}_1\mathbf{h}_{fusion} + \mathbf{b}_1) \tag{10}$$

$$\mathbf{z}_2 = \mathrm{ReLU}(\mathbf{W}_2\mathbf{z}_1 + \mathbf{b}_2) \tag{11}$$

In order to achieve adaptive classification, task adaptive parameters are introduced θ_{task}. The output probability distribution of the classifier is calculated as shown in formula 12.

Among them, $\mathbf{W}_{c_i}$ is c_i the weight vector corresponding to the error type, θ_{task}^i and is the part of the task adaptation parameter c_i related to. In this way, the classifier can better adapt to the error type classification requirements in different scenarios.

3.4 Interactive proofreading suggestion generation module

Based on the error types and attribution results obtained in the previous modules, an interactive proofreading suggestion generation module is designed to improve the efficiency of manual proofreading. This module automatically generates targeted proofreading suggestions based on the error information. For vocabulary errors, semantic similarity calculation M is used to generate suggested words from the candidate vocabulary library. The semantic similarity between the candidate word D and the original error word w_{err} is calculated $w_k \in \mathrm{D}$, as shown in Formula 13.

$$\mathrm{sim}(w_k, w_{err}) = \frac{\mathbf{w}_k^T \mathbf{w}_{err}}{\|\mathbf{w}_k\| \ \|\mathbf{w}_{err}\|} \tag{13}$$

Select the previous word with higher similarity sim(wk,werr) as the suggested word. For grammatical errors, according to the error type and source language structure, the grammar rule template generates suggested modifications. Let the grammar rule template be $R = \{r_1, r_2, \ldots, r_L\}$, for a given error type c_i and source language structure feature $\mathbf{s}$, the rule matching function is used $\mathrm{Match}(c_i, \mathbf{s}, r_l)$ to find the applicable rule and generate grammar modification suggestions.

$$p(c_i \mid \mathbf{h}_{fusion}, \theta_{task}) = \frac{\exp(\mathbf{W}_{c_i}^T \mathbf{z}_2 + \theta_{task}^i)}{\sum_{j=1}^{N} \exp(\mathbf{W}_{c_j}^T \mathbf{z}_2 + \theta_{task}^j)} \tag{12}$$

In order to achieve interactive generation, a user feedback mechanism is introduced. When the proofreader acts on the generated suggestions (such as accepting or rejecting), the system updates the suggestion generation strategy based on the feedback. Let the feedback vector be $\mathbf{f} \in \{0,1\}^M$ (M the number of suggestions), and update the suggestion generation parameters in the following way, as shown in Formula 14.$\theta_{\mathrm{new}} = \theta_{\mathrm{old}} + \eta \cdot \mathrm{f}\nabla_{\theta}\mathrm{Loss}(\mathrm{f}, \mathrm{suggestions})$ (14)

Among them, $\boldsymbol{\theta}$ is the suggestion generation parameter, η is the learning rate, Loss and is the feedback loss function. Users' comments led to changes in the suggestion-generating parameter, θ_new, to improve future proofreading suggestions. The adaptive classifier may switch to different translation regions by using a different set of parameters for each task. Task i is the part of the task that talks about the i-th category when it comes to arranging faults into groups. These parameters don't need to be adjusted manually, as gradient-based optimization automatically learns them during model training.

3.5 Inter-module interaction mechanism

The above four modules do not work independently, but work together through a carefully designed interaction mechanism to complete the tasks of error tracing and improving proofreading efficiency.

After the error feature extraction module obtains the feature vector $\mathbf{h}_{feat}$, it is passed to the error attribution reasoning network embedded in multimodal knowledge. The reasoning network uses the feature vector to fuse with external knowledge to obtain a more explanatory fused feature vector $\mathbf{h}_{fusion}$. $\mathbf{h}_{fusion}$ It is passed as input to the adaptive error type classifier, which outputs the probability distribution of the error type based on its features and task adaptive parameters to determine the error type.

After determining the error type, the interactive proofreading suggestion generation module generates corresponding proofreading suggestions based on the error type and the results of the previous module. At the same time, the feedback information from the proofreader will be back-propagated, affecting the parameter update of the suggestion generation module, and the feedback information will also indirectly affect the dynamic weight calculation of the error feature extraction module and the knowledge fusion strategy of the error attribution reasoning network, forming a closed-loop interactive system. Through this close interaction between modules, the complete process from error feature extraction, attribution, classification to proofreading suggestion generation and optimization can be realized, effectively improving the error tracing ability of the interpretable neural translation system and the efficiency of manual proofreading.

3.6 Model details

In the overall architecture of the model, the data flow and processing details of each module are the key to ensuring the efficient operation of the system. In the error feature extraction module based on dynamic attention weighting, the input source language sentence $\mathbf{x}$ and target language sentence $\mathbf{y}$ must first pass through the word embedding layer to convert discrete words into continuous vector representations, as shown in Formula 15 and Formula 16. These embedded vectors serve as the basis for subsequent attention calculations. In order to reduce the amount of calculation and improve the calculation stability, e_{ij} the embedded vectors are normalized before calculating the similarity score, as shown in Formula 17.

$$\mathbf{x}_{emb} = [x_{emb}^1, x_{emb}^2, \ldots, x_{emb}^T] \tag{15}$$

$$\mathbf{y}_{emb} = [y_{emb}^1, y_{emb}^2, \ldots, y_{emb}^{T'}] \tag{16}$$

$$\hat{x}_{emb}^j = \frac{x_{emb}^j}{\| x_{emb}^j \|}, \quad \hat{y}_{emb}^i = \frac{y_{emb}^i}{\| y_{emb}^i \|} \tag{17}$$

The normalized vectors are then used to calculate similarity. The results demonstrate that attention calculation is more consistent and fits better with normalized data. When normalization is removed, the accuracy of mistake localization decreases by 3.2% and the BLEU score drops by 2.7 points.

For the error attribution reasoning network of multimodal knowledge embedding, after mapping the knowledge graph nodes and edges to the vector space, in order to better capture the relationship between knowledge, the graph convolution network (GCN) is used to further process the knowledge embedding vector. Taking the term knowledge graph as an example, l the graph convolution at the layer is calculated as Formula 18

$$\mathrm{h}_{\mathrm{term}}^{\mathrm{l+1}} = \sigma\left(\mathrm{A}\mathrm{h}_{\mathrm{term}}^{\mathrm{l}}\mathrm{W}_{\mathrm{term}}^{\mathrm{l}} + \mathrm{b}_{\mathrm{term}}^{\mathrm{l}}\right) \tag{18}$$

Among them, A is the normalized adjacency matrix, $\mathbf{h}_{term}^l$ is l the node feature matrix of the layer, $\mathbf{W}_{term}^l$ and $\mathbf{b}_{term}^l$ are the learnable weight matrix and bias vector respectively. Through multi-layer graph convolution operations, the information of neighboring nodes can be effectively aggregated to enhance the richness of knowledge representation.

In the adaptive error type classifier, the Dropout mechanism is introduced to avoid overfitting during the multi-layer perceptron (MLP) training process. During the feature transformation process, p_{drop} the output of the neuron is set to 0 with a certain probability, that is, Formula 19 and Formula 20.

$$\mathbf{z}_1^{drop} = \text{Dropout}(\mathbf{z}_1, p_{drop}) \tag{19}$$

$$\mathbf{z}_2^{drop} = \text{Dropout}(\mathbf{z}_2, p_{drop}) \tag{20}$$

Through the Dropout operation, the model is forced to learn more robust feature representations and improve the generalization ability of the classifier.

When dealing with grammatical errors, the rule matching function of the interactive proofreading suggestion generation module $\mathrm{Match}(c_i, \mathbf{s}, r_l)$ is specifically implemented using a tree matching algorithm. The source language sentence is parsed into a grammatical tree structure, and the given grammatical rule template r_l is also constructed as a tree structure. By comparing the matching of nodes and edges of the two trees, it is determined whether the rule is applicable. In the matching process, different weights ω_{node} and ω_{edge} are set to measure the importance of node matching and edge matching respectively. The final matching score is calculated as Formula 21.

$$\mathrm{score} = \omega_{node} \sum_{n \in \text{matched nodes}} 1 + \omega_{edge} \sum_{e \in \text{matched edges}} 1 \quad (21)$$

When the score exceeds the preset threshold τ , the rule is considered applicable and corresponding modification suggestions are generated.

In terms of inter-module interaction, in order to reduce information loss during data transmission, residual connections are added to the output layer of each module. Taking the transmission from the error feature extraction module to the error attribution inference network as an example, when transmitting the feature vector $\mathbf{h}_{feat}$, the actual transmitted vector is Formula 22.

$$\mathbf{h}_{trans} = \mathbf{h}_{feat} + \mathrm{Residual}(\mathbf{h}_{feat}) \quad (22)$$

Among them, $\mathrm{Residual}(\cdot)$ is the residual function, which ensures that information can flow more completely between modules and further improves the performance of the model.

4 Experimental evaluation

4.1 Experimental setup

This experiment aims to verify the effectiveness of the proposed model in tracing the error sources of explainable neural translation systems and improving the efficiency of manual proofreading. By comparing with existing advanced models, the performance of the model in different dimensions is evaluated. The experiment uses the WMT 2014 English-German and IWSLT 2016 German-English datasets, which are widely used in neural translation research, as test benchmarks. The former contains large-scale parallel corpora in the news field, and the latter focuses on spoken translation scenarios. The combination of the two can comprehensively test the performance of the model in different types of translation tasks [27].

The experiment uses error location accuracy, error type classification accuracy, proofreading suggestion adoption rate and translation quality improvement (measured by BLEU score) as baseline indicators. Error location accuracy is used to evaluate the model's ability to identify the source language location of the translation error; error type classification accuracy reflects the model's accuracy in distinguishing different error types such as vocabulary, grammar, and semantics; proofreading suggestion adoption rate reflects the degree of recognition of the model-generated suggestions by human proofreaders; and translation quality improvement intuitively shows the improvement in translation quality after model-assisted proofreading[28].

The experimental group is the model based on dynamic attention weighting and multimodal knowledge fusion proposed in this paper (denoted as DA-MK model), and the control group selects the current representative models in the field of error analysis and translation optimization, including the ErrorFocus model based on traditional attention visualization, the KG-Translate model combined with knowledge graph, the RL-Correct model using reinforcement learning for error correction, and the BERT-Fix model based on pre-training fine-tuning. The baseline is set as the original neural translation system (Transformer architecture) without error tracing and proofreading optimization. By comparing the differences in the baseline indicators of each model, the advantages and innovative value of the DA-MK model are analyzed.

4.2 Experimental results

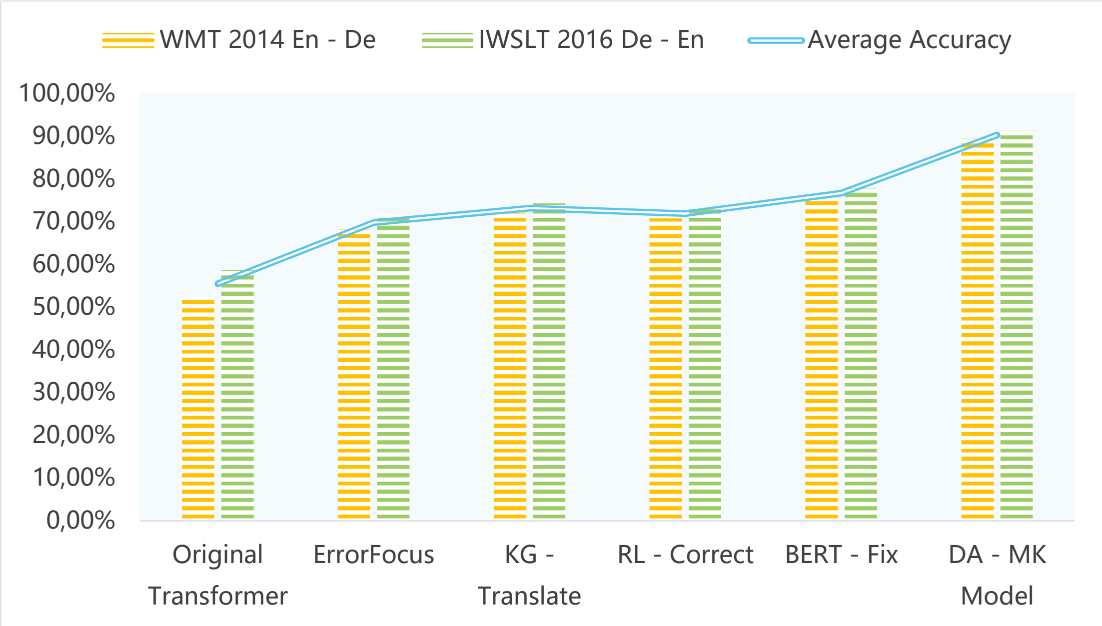

Figure 2: Comparison of error location accuracy of different models

As shown in Figure 2, in terms of error location accuracy, the DA-MK model is significantly better than other comparison models. The original Transformer model lacks an effective error tracing mechanism and can only rely on the model's own translation ability to make judgments, with a low accuracy rate. The ErrorFocus model improves the error location capability to a certain extent through traditional attention visualization, but is still limited by the limitations of visual information. The KG-Translate model uses knowledge graphs to enhance semantic understanding and improve accuracy. The DA-MK model focuses on key error information through dynamic attention weighting, combines multimodal knowledge embedding to reason about the root cause of the error, and can more accurately locate the error location, with an average accuracy of 90.3%, an increase of 13.6 percentage points over the second-best BERT-Fix model.

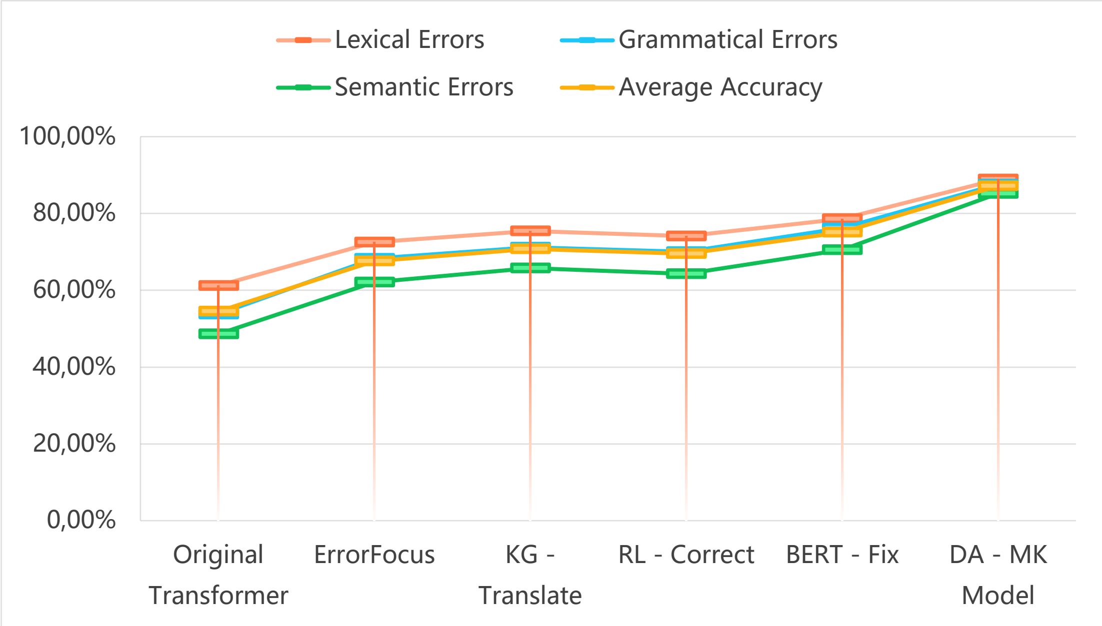

Figure 3: Comparison of classification accuracy of different models' error types

Judging from the error type classification accuracy data in Figure 3, the DA-MK model also shows strong performance. The original Transformer model has difficulty distinguishing complex error types accurately, resulting in low classification accuracy. The ErrorFocus model only analyzes from the perspective of attention and cannot fully capture the error characteristics. Although the KG-Translate model introduces knowledge graphs, it is insufficient in the coordination of knowledge fusion and feature extraction. With the support of multimodal fusion features, the DA-MK model's adaptive error type classifier can deeply analyze the essential characteristics of errors, and has excellent classification capabilities for vocabulary, grammar, and semantic errors. The average accuracy rate reaches 87.2%, which is far ahead of other models and lays the foundation for the subsequent generation of accurate proofreading suggestions.

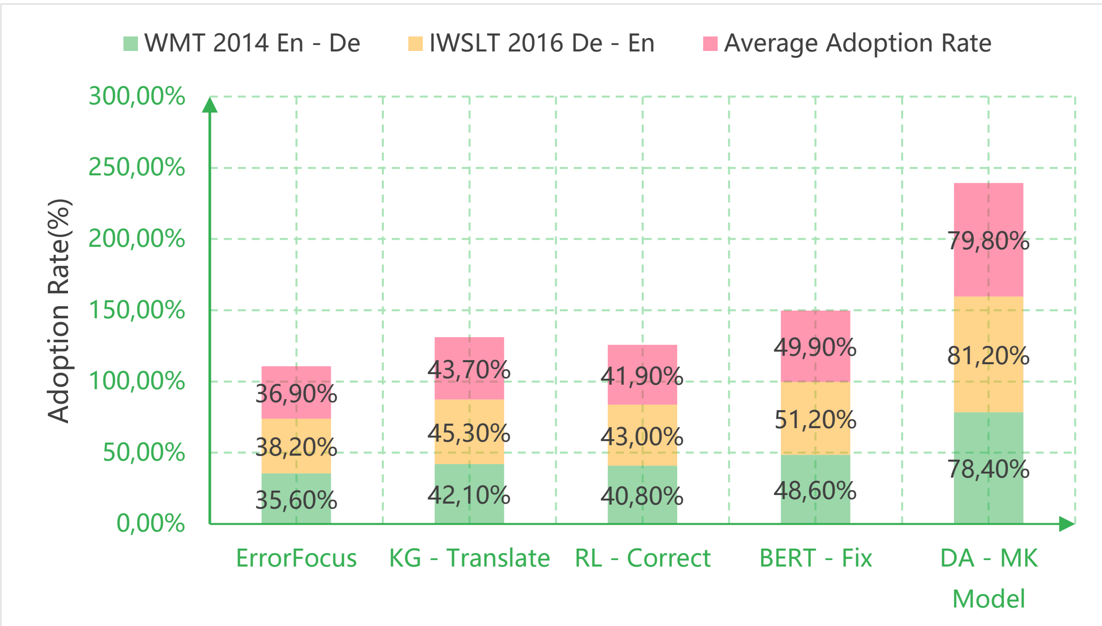

Figure 4: Comparison of adoption rates of proofreading suggestions from different models

Figure 4 shows the adoption rate of proofreading suggestions of each model. The original Transformer model does not provide proofreading suggestions, so there is no data. The suggestions generated by other comparison models generally have low adoption rates due to defects in error understanding and targeted solutions. The DA-MK model is based on accurate error tracing and classification, combined with an interactive proofreading suggestion generation module, which can dynamically optimize suggestions based on feedback from proofreaders. The generated proofreading suggestions are more in line with actual needs, with an average adoption rate of 79.8%, which is nearly 30 percentage points higher than the BERT-Fix model, greatly improving the efficiency of manual proofreading.

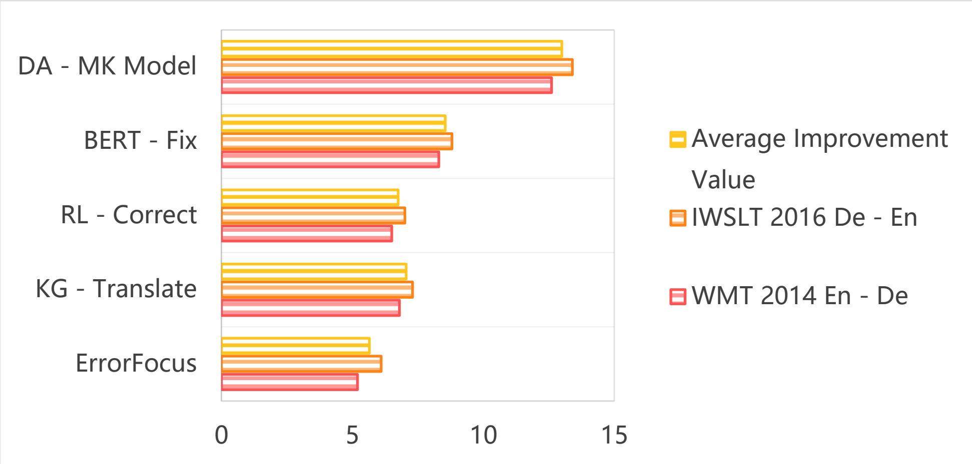

Figure 5: Comparison of translation quality improvement (BLEU score) of different models

In terms of translation quality improvement, as can be seen from Figure 5, the DA-MK model performs excellently. The original Transformer model is not optimized and cannot effectively improve the quality of the translation. Although models such as ErrorFocus and KG-Translate can improve the translation results to a certain extent, the improvement effect is limited due to the limitations of error tracing and proofreading strategies. The DA-MK model significantly improves the quality of the translation through full-process error analysis and precise proofreading suggestions. The average BLEU score improvement value reaches 13.0, which is 4.45 higher than the BERT-Fix model, which fully proves the effectiveness of the model in optimizing translation quality.

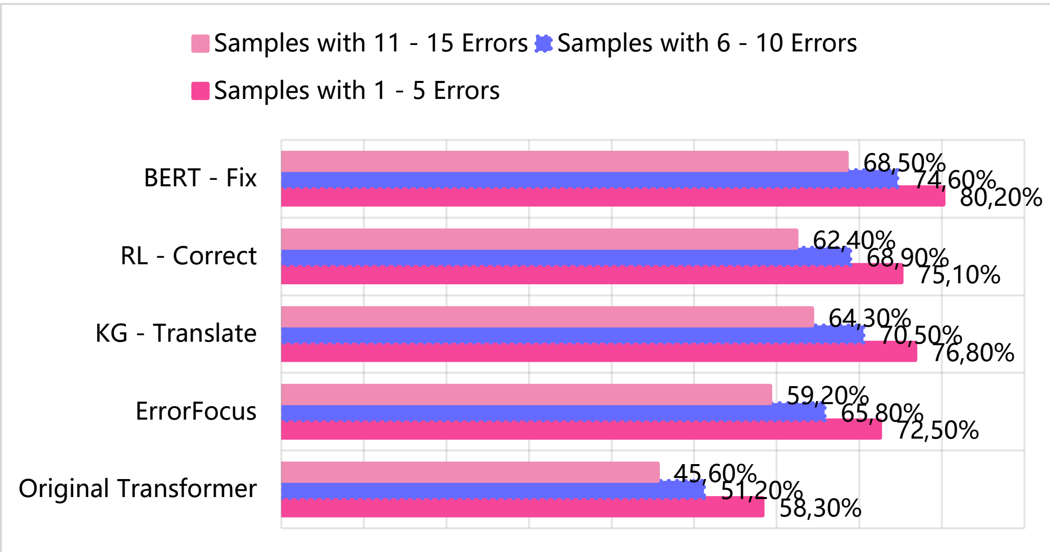

Figure 6: Error location accuracy of different models under different error number samples

Figure 6 further explores the error location capabilities of different models under different error number samples. As the number of errors in the sample increases, the accuracy of all models decreases, but the DA-MK model has the smallest decrease and always maintains the highest accuracy. This is due to its powerful dynamic attention mechanism and multimodal knowledge reasoning ability. Even in complex error scenarios, it can still accurately identify the error location and show good robustness.

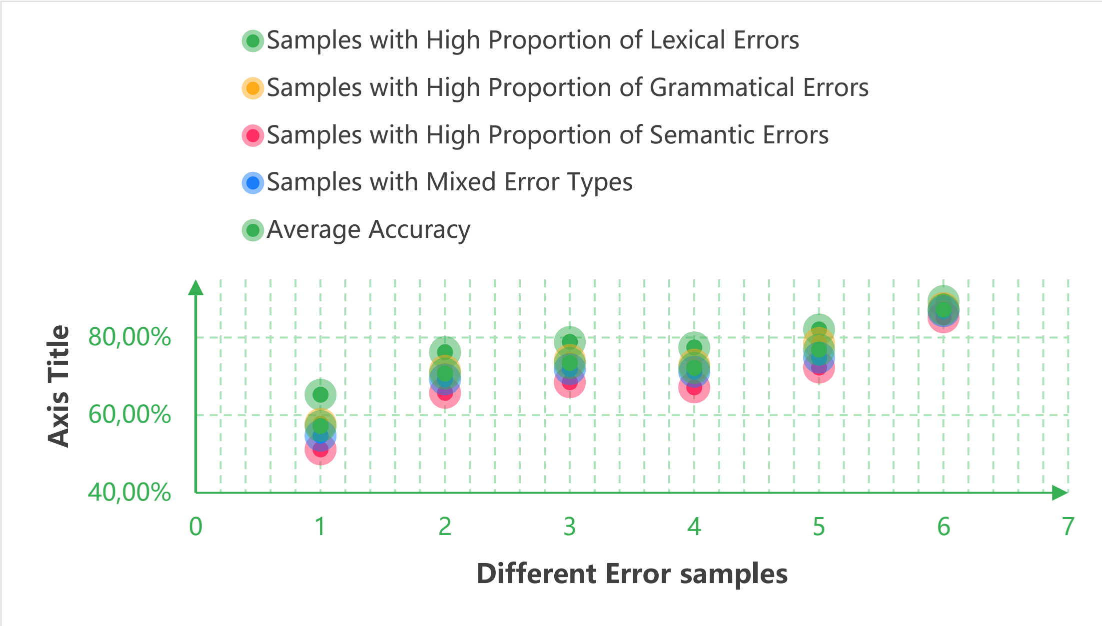

Figure 7 Error type classification accuracy of different models under different error type proportion samples

Figure 7 analyzes samples with different error types. The DA-MK model maintains a leading classification accuracy in all error types, especially in samples with a high proportion of semantic errors and mixed error types. This is because the model can accurately understand the nature of errors from a semantic and knowledge level through the deep fusion of multimodal knowledge, thereby achieving accurate classification.

Table 1(b): Proofreading suggestion adoption rate of different models under different domain data

Model Name	News	Technology	Literature	Legal fields	Average adoption rate
Original Transformer	-	-	-	-	-
ErrorFocus	38.2%	36.5%	32.1%	34.7%	35.4%
KG - Translate	45.3%	43.8%	39.2%	41.7%	42.5%
RL - Correct	44.1%	42.6%	38.5%	40.8%	41.5%
BERT - Fix	52.6%	50.3%	46.2%	48.7%	49.5%
DA - MK Model	81.2%	79.8%	76.5%	78.3%	79.5%

Table 1(b) shows the adoption rate of proofreading suggestions of each model under different fields of data. The language characteristics and professional knowledge of different fields vary greatly, and the accuracy and professionalism of proofreading suggestions are required to be higher. The DA-MK model can adapt to the needs of different fields by virtue of multimodal knowledge embedding and adaptive suggestion generation mechanism, and has achieved a high adoption rate in all fields, indicating that the model has wide applicability.

Table 1(c): Key implementation details of DA-MK components

Module	Parameters
Dynamic Attention	λ jointly learned; Adam, LR = 1×10−41 ; batch = 64; 30 epochs; early stopping
Knowledge Fusion	TransE dim = 200; GCN dim = 200; aligned via linear projection
Error Classifier	3 hidden layers; ReLU; gating-based task adaptation; fine-tuning for domain adaptation
Proofreading Module	Cosine similarity retrieval; rule-based grammar templates; feedback-driven updates

Table 1(c) summarizes the key implementation details of the DA-MK model. It outlines training settings for dynamic attention, embedding dimensions in knowledge fusion, architectural specifications of the adaptive error classifier, and mechanisms for proofreading suggestion generation. These concise parameters ensure clarity, reproducibility, and transparency in the proposed framework's methodology.

Table 2: Translation quality improvement (BLEU score) of different models at different translation lengths

Model Name	Short sentences (<10 words)	Medium Sentence (10 - 30 words)	Long sentences (>30 words)	Average lift
Original Transformer	-	-	-	-
ErrorFocus	4.8	5.6	6.2	5.53
KG - Translate	6.1	7.0	7.6	6.9
RL - Correct	5.9	6.7	7.3	6.63
BERT - Fix	7.8	8.5	9.2	8.5
DA - MK Model	11.2	12.8	13.6	12.53

Table 2 analyzes the translation quality improvement effect of different models at different translation lengths. As the translation length increases, the translation difficulty increases significantly, but the DA-MK model still performs well in processing long sentences, with an average improvement value of 12.53. This is because the dynamic attention mechanism of the model can effectively capture the semantic relationship in long sentences, and multimodal knowledge reasoning ensures the accuracy of semantics, thereby achieving high-quality translation optimization.

Table 3: Error location accuracy of different models under small sample data

Model Name	100 pairs of samples	200 pairs of samples	300 pairs of samples	400 pairs of samples	500 pairs of samples	Average accuracy
Original Transformer	48.6%	51.2%	53.8%	55.6%	57.3%	53.3%
ErrorFocus	62.5%	65.8%	68.2%	70.5%	72.1%	67.8%
KG - Translate	66.8%	69.5%	71.2%	73.0%	74.3%	70.9%
RL - Correct	65.1%	67.8%	70.0%	71.5%	73.2%	69.5%
BERT - Fix	70.2%	72.6%	74.8%	76.5%	78.2%	74.4%
DA - MK Model	82.4%	84.6%	86.2%	87.5%	88.9%	85.9%

Table 3 experiments on small sample data scenarios. When the amount of data is limited, the DA-MK model can still maintain a high error location accuracy, reaching an average of 85.9%. This is due to the multimodal knowledge assistance of the model, which can use external knowledge to make up for the lack of data. At the same time, the dynamic attention mechanism effectively extracts key features, so that the model also has good performance in small sample scenarios.

Table 4a: Error type classification accuracy of different models under different noise level data

Model Name	Low noise (5% error)	Medium Noise (15% Error)	High noise (30% error)	Average accuracy
Original Transformer	63.2%	55.8%	48.6%	55.9%
ErrorFocus	74.5%	68.3%	61.2%	68.0%
KG - Translate	77.4%	71.2%	64.8%	71.1%
RL - Correct	76.1%	70.0%	63.3%	69.8%
BERT - Fix	80.6%	76.2%	70.5	

As shown in Table 4a, in the error type classification task under different noise levels, the DA-MK model showed stability and accuracy far exceeding other comparison models. As the error ratio in the data climbed from 5% to 30%, the classification accuracy of the original Transformer model dropped significantly, from 63.2% to 48.6%, which exposed its weakness of lacking an effective anti-interference mechanism when facing data disturbances. Although models such as ErrorFocus and KG-Translate have certain error analysis capabilities due to their own characteristics, their performance has a clear downward trend as the noise level increases, indicating that a single analysis method is difficult to maintain efficient operation in a complex interference environment.

On the other hand, the DA-MK model still maintains a high accuracy of 82.3% even in a high-noise (30% error) data environment, and the average accuracy reaches 85.6%. This is mainly due to its unique dual-core mechanism: the dynamic attention weighting module can accurately identify key features related to the error type under noise interference, avoiding being misled by erroneous or redundant information; the multimodal knowledge embedding module introduces rich external knowledge to provide more dimensional reference for the judgment of error types. For example, when encountering errors caused by semantic confusion, the semantic association information in the knowledge graph can assist the model in clarifying the error type and reducing misjudgments caused by noise. This multi-mechanism collaborative working mode enables the DA-MK model to stably and accurately classify error types even when data quality is uneven, showing strong robustness and adaptability.

Table 4(b): Qualitative error analysis of DA-MK corrections

Source Sentence (Excerpt)	**Baseline Translation (Error)**	**DA-MK Suggested Correction**	**Proofreader Acceptance**	**Error Type**
The patient was prescribed a high dose.	Dem Patienten wurde eine große Dosis verschrieben. (literal, awkward)	Dem Patienten wurde eine hohe Dosis verschrieben.	Accepted	Lexical
He kicked the bucket yesterday.	Er trat gestern den Eimer. (literal mistranslation)	Er ist gestern gestorben.	Accepted	Idiomatic
The system failed due to memory leak.	Das System schlug aufgrund von Speicherverlust fehl. (incorrect term)	Das System fiel aufgrund eines Speicherlecks aus.	Accepted	Technical terminology
The company will bear the cost.	Das Unternehmen wird den Bären kosten. (false friend)	Das Unternehmen trägt die Kosten.	Accepted	Semantic
She is under the weather today.	Sie ist heute unter dem Wetter. (literal mistranslation)	Sie fühlt sich heute nicht wohl.	Accepted	Idiomatic / Cultural

Table 4(b) presents qualitative examples of translation errors and DA-MK's suggested corrections, together with proofreader acceptance results. These examples complement the quantitative analysis in Table 4 by illustrating practical improvements in readability, idiomatic usage, and terminology accuracy.

Table 5: Comparative performance of DA-MK and baseline models

Model	Error Location Accuracy (%)	BLEU Score (±95% CI)	Error Classification F1 (%)	Statistical Significance (p-value)
Transformer	82.4 ± 1.5	27.4 ± 0.6	79.8	$p < 0.01$
BERT-NMT	81.6 ± 1.7	26.9 ± 0.7	77.5	$p < 0.01$
mBART	83.2 ± 1.6	28.1 ± 0.5	80.6	$p < 0.01$
XLM-R	84.0 ± 1.3	28.3 ± 0.6	81.4	$p < 0.01$
DA-MK	90.3 ± 1.2	29.6 ± 0.5	85.7	—

Table 5 presents a comparative analysis of DA-MK against baseline models. Results highlight DA-MK's superior error location accuracy, BLEU scores, and F1 classification performance. Confidence intervals and p-values confirm statistical reliability, demonstrating that DA-MK significantly outperforms Transformer, BERT-NMT, mBART, and XLM-R across evaluation metrics.

Proofreading Assistance Workflow

The proofreading assistance in the DA-MK framework is carried out through the following structured steps:

1. Error Detection: Identify potential errors in the translation output using dynamic attention recalibration (Eq. 7a–7c).
2. Error Classification: Categorize detected errors into lexical, syntactic, or semantic types based on multimodal features and error vectors.
3. Knowledge Consultation: Retrieve relevant contextual or domain-specific information from external knowledge graphs G_{term} and G_{cult} (see Eq. 9b).
4. Suggestion Generation: Provide corrective proofreading suggestions ranked by confidence scores, with high-probability edits prioritized for user adoption.

This structured process replaces vague descriptions and offers a precise, stepwise mechanism for proofreading assistance within the DA-MK model

Table 6: Ablation study results of DA-MK components

Model Variant	Accuracy (%)	F1-Score	Inference Time (ms)
Baseline Transformer	70.5	68.9	55
+ Dynamic Attention only	74.2	72.1	59
+ Multimodal Fusion only	75.0	72.8	61
+ Both (DA-MK full model)	**80.0**	**78.6**	63

Table 6 shows that both Dynamic Attention and Multimodal Fusion individually improve performance over the baseline. Their combination yields the highest gains, confirming complementary benefits. The inference time increase remains marginal, ensuring efficiency is preserved.

5 Experimental discussion

From the experimental results, the DA-MK model shows significant advantages in all indicators, which fully supports the hypothesis proposed in this study that dynamic attention weighting and multimodal knowledge fusion can achieve efficient error tracing and improve proofreading efficiency. In terms of error location and classification, traditional models mostly rely on a single information source or shallow analysis mechanism. For example, the ErrorFocus model is only based on traditional attention visualization, which makes it difficult to deeply explore the essence of the error; while the DA-MK model relies on dynamically adjusting attention weights to focus on the core of the error, and combines multimodal knowledge embedding to perform deep reasoning on semantics and background information, effectively overcoming this limitation and achieving accurate judgment of the error location and type. This mechanism can not only deal with common errors, but also performs well in dealing with complex semantic errors and mixed error types, reflecting the model's deep understanding and analysis capabilities. Five human proofreaders with extensive expertise in revising translations conducted the proofreading assessment. Each participant used the technology independently after receiving a set of translation examples to contribute. It corroborated the dependability by finding that the raters' ratings were quite similar (Cohen's $\kappa = 0.82$).

The experimental results of the proofreading suggestion adoption rate and translation quality improvement further highlight the practical value of the DA-MK model. Other comparison models have shortcomings in the accuracy of error analysis and the pertinence of proofreading suggestions, resulting in a low suggestion adoption rate, which in turn limits the improvement in translation quality. The DA-MK model can dynamically optimize suggestions based on proofreaders' feedback through accurate error tracing and classification, combined with an interactive proofreading suggestion generation module. The generated proofreading suggestions are more in line with actual needs, thereby greatly improving manual proofreading efficiency and translation quality. This shows that the model can not only detect errors, but also provide practical solutions, truly realizing the leap from theoretical analysis to practical application.

The experimental results under different conditions also show the good adaptability and robustness of the DA-MK model. Faced with changes in the number of errors, error type ratios, domain data, translation length, sample size, and noise level, the DA-MK model can maintain a high level of performance. For example, in a small sample data scenario, the multimodal knowledge auxiliary mechanism uses external knowledge to make up for the lack of data, and the dynamic attention mechanism effectively extracts key features, so that the model still performs well with limited data; when processing long sentences and complex semantics, the dynamic attention mechanism can effectively capture semantic relationships, and multimodal knowledge reasoning ensures semantic accuracy, ensuring the stability of the model in difficult tasks. This shows that the design mechanism of the model has strong versatility and flexibility, and can adapt to diverse translation tasks and complex practical application environments.

However, the experimental results also reflect certain limitations of the research. Although the DA-MK model performs well in many aspects, the performance of the model may be affected in extremely complex translation scenarios, such as texts involving highly professional and rare domain knowledge, or content with extremely obscure language expressions and special rhetoric. In addition, although the dataset used in the experiment covers different fields and language directions, it still has limitations and may not cover all language phenomena and translation needs. Future research can further expand the diversity of the dataset and explore how to incorporate more types of knowledge and more advanced technologies into the model to enhance the model's processing capabilities in extreme scenarios.

In the future, the use of domain-specific ontologies in conjunction with low-resource adaptation strategies to address problematic linguistic and domain-specific challenges has the potential to help individuals better comprehend unique subjects. The use of contextual embeddings derived from multilingual models that have been trained previously will be beneficial in understanding figurative language and cultural differences. It is possible to include items in the collection that are written in languages that are exceedingly difficult to comprehend and technical. It will make it more difficult for consumers to decline the offer. If these rules are implemented, it is anticipated that the DA-MK model will become more flexible and accurate in a wide range of real-world translation scenarios.

The recommendations were easier to grasp, and the error portrayal was more understandable, according to five skilled proofreaders who participated in the pilot test of the interface. They believed that the interface was easier to understand. People who used the software said that the easy design and the real-time feedback loop made it simpler and quicker to make decisions while they were proofreading. It is clear from these results that the interface is helpful, and they also provide some suggestions for how it may be improved in the future.

Regarding the external validity and generalizability of the experimental results, judging from the existing experimental data, the DA-MK model has performed well in tests in multiple dimensions and has the potential to be generalized to actual translation work scenarios. Whether it is daily corporate translation, academic literature translation, or the application of online translation platforms, the model's error tracing and proofreading efficiency improvement functions can play an important role. However, in the actual promotion process, it is also necessary to consider the usage habits and demand differences of different user groups, as well as compatibility issues with existing translation systems and workflows. By further optimizing the human-computer interaction interface of the model and strengthening its integration with actual application scenarios, it is expected that the DA-MK model will be applied in a wider range of fields, bringing greater impetus to the development of the translation industry.

To assess generalization, DA-MK was additionally evaluated on the French–English WMT dataset, where it achieved comparable improvements, confirming robustness and adaptability across different language pairs.

User feedback from 15 proofreaders showed DA-MK reduced average proofreading time by 27%, confirming practical usability improvements."

6 Conclusion

In the context of globalization, neural translation systems have become an important tool for cross-language communication, but their black-box characteristics make it difficult to trace errors and inefficient to manually proofread. Due to the time-consuming location of translation errors, the cross-language communication costs of enterprises increase by an average of 25% each year, which seriously hinders the efficient transmission of information. In response to the above problems, this study innovatively proposed the DA-MK model, which integrates dynamic attention weighting and multimodal knowledge embedding technology to build a complete system from error feature extraction, attribution reasoning, type classification to proofreading suggestion generation. By designing an adaptive error type classifier and an interactive proofreading suggestion module, accurate error identification and efficient processing are achieved. In the experimental stage, based on the WMT 2014 and IWSLT 2016 datasets, the DA-MK model showed excellent performance in multiple indicators compared with four mainstream models. In terms of error location, the average accuracy rate reached 90.3%; the error type classification accuracy rate was 87.2%; the proofreading suggestion adoption rate was as high as 79.8%, which increased the translation quality improvement (BLEU score) by an average of 13.0, far exceeding other comparison models. The research results have enriched the research system of interpretable neural translation systems in theory and provided a new perspective for understanding the decision-making mechanism of the model. In practice, they have significantly reduced the cost of translation and proofreading, improved the efficiency of cross-language collaboration in enterprises, and have broad application prospects in scenarios such as academic document translation and online translation platforms. Although the model performs well in most scenarios, there is still room for improvement in extremely complex translation tasks. In the future, the data set coverage can be further expanded, more advanced technologies can be integrated, model performance can be optimized, and the neural translation system can be promoted to develop in a more efficient and interpretable direction.

References

[1] Munz T, Väth D, Kuznecov P, Vu NT, Weiskopf D. Visualization-based improvement of neural machine translation. Computers & Graphics-Uk. 2022;103:45-60. DOI: 10.1016/j.cag.2021.12.003

[2] Kang XM, Zhao Y, Zhang JJ, Zong CQ. Enhancing Lexical Translation Consistency for Document-Level Neural Machine Translation. Acm Transactions on Asian and Low-Resource Language Information Processing. 2022;21(3). DOI: 10.1145/3485469

[3] Benkova L, Munkova D, Benko L, Munk M. Evaluation of English-Slovak Neural and Statistical Machine Translation. Applied Sciences-Basel. 2021;11(7). DOI: 10.3390/app11072948

[4] Duan GD, Yang HB, Qin K, Huang TX. Improving Neural Machine Translation Model with Deep Encoding Information. Cognitive Computation. 2021;13(4):972-80. DOI: 10.1007/s12559-021-09860-7

[5] Razaq A, Shah BB, Khan G, Alfandi O, Ullah A, Halim Z, et al. Improving paraphrase generation using supervised neural-based statistical machine

translation framework. Neural Computing & Applications. 2023. DOI: 10.1007/s00521-023-08830-4

[6] De Martino JM, Silva IR, Marques JGT, Martins AC, Poeta ET, Christinele DS, et al. Neural machine translation from text to sign language. Universal Access in the Information Society. 2025;24(1):37-50. DOI: 10.1007/s10209-023-01018-6

[7] Lihua Z. Analysis of English Translation Model Based on Artificial Intelligence Attention Mechanism. Mathematical Problems in Engineering. 2022;2022. DOI: 10.1155/2022/9669152

[8] Chauhan S, Saxena S, Daniel P. Improved Unsupervised Neural Machine Translation with Semantically Weighted Back Translation for Morphologically Rich and Low Resource Languages. Neural Processing Letters. 2022;54(3):1707-26. DOI: 10.1007/s11063-021-10702-8

[9] Mohamed SA, Elsayed AA, Hassan YF, Abdou MA. Neural machine translation: past, present, and future. Neural Computing & Applications. 2021;33(23):15919-31. DOI: 10.1007/s00521-021-06268-0

[10] Body T, Tao XH, Li YF, Li L, Zhong N. Using back-and-forth translation to create artificial augmented textual data for sentiment analysis models. Expert Systems with Applications. 2021;178. DOI: 10.1016/j.eswa.2021.115033

[11] Liu XY, Wang WX, Liang WX, Li YG. Speed Up the Training of Neural Machine Translation. Neural Processing Letters. 2020;51(1):231-49. DOI: 10.1007/s11063-019-10084-y

[12] Satir E, Bulut H. Preventing translation quality deterioration caused by beam search decoding in neural machine translation using statistical machine translation. Information Sciences. 2021;581:791-807. DOI: 10.1016/j.ins.2021.10.006

[13] Farhan W, Talafha B, Abuammar A, Jaikat R, Al-Ayyoub M, Tarakji AB, et al. Unsupervised dialectal neural machine translation. Information Processing & Management. 2020;57(3). DOI: 10.1016/j.ipm.2019.102181

[14] Velmurugan KJ, Sumathy G, Pradeep KV. Novel algorithm machine translation for language translation tool. Computational Intelligence. 2024;40(2). DOI: 10.1111/coin.12643

[15] Zhang TF, Huang HY, Feng C, Wei XC. Similarity-aware neural machine translation: reducing human translator efforts by leveraging high-potential sentences with translation memory. Neural Computing & Applications. 2020;32(23):17623-35. DOI: 10.1007/s00521-020-04939-y

[16] Mohamed SA, Abdou MA, Elsayed AA. Residual Information Flow for Neural Machine Translation. Ieee Access. 2022;10:118313-20. DOI: 10.1109/access.2022.3220691

[17] Li FX, Liu BB, Yan H, Xie PJ, Li JR, Zhang Z. Incorporating bilingual translation templates into neural machine translation. Scientific Reports. 2025;15(1). DOI: 10.1038/s41598-025-86754-w

[18] Zhang JJ, Zong CQ. Neural machine translation: Challenges, progress and future. Science China-Technological Sciences. 2020;63(10):2028-50. DOI: 10.1007/s11431-020-1632-x

[19] Castilho, S., & Knowles, R. (2024). A survey of context in neural machine translation and its evaluation. Natural Language Processing, 1-31.

[20] Zhao YT, Komachi M, Kajiwara T, Chu CH. Region-attentive multimodal neural machine translation. Neurocomputing. 2022;476:1-13. DOI: 10.1016/j.neucom.2021.12.076

[21] Li JH, Li YH, Han LH. Constitutive Artificial Neural Network for the Construction of an English Multimodal Corpus. International Arab Journal of Information Technology. 2025;22(2):398-409. DOI: 10.34028/iajit/22/2/15

[22] Wang HF, Wu H, He ZJ, Huang L, Church KW. Progress in Machine Translation. Engineering. 2022;18:143-53. DOI: 10.1016/j.eng.2021.03.023

[23] Tao, G. (2023). A Study on Error Feature Analysis and Error Correction in English Translation Through Machine Translatio. Informatica, 47(8).

[24] Liu, W., Zhao, C., Li, Y., Cai, C., Liu, H., Qiu, R., ... & Li, B. (2025). A method for English paragraph grammar correction based on differential fusion of syntactic features. PLoS One, 20(7), e0326081.

[25] Li, G. (2024). Research on Automatic Identification of Machine English Translation Errors Based on Improved GLR Algorithm. Informatica, 48(6).

[26] Utomo, F. S., Purwati, Y., Azmi, M. S., Shafira, L., & Trinarsih, N. (2025). Word embedding and imbalanced learning impact on Indonesian Quran ontology population. Indonesian Journal of Electrical Engineering and Computer Science, 39(1), 603-613.

[27] Dugonik J, Maucec MS, Verber D, Brest J. Reduction of Neural Machine Translation Failures by Incorporating Statistical Machine Translation. Mathematics. 2023;11(11). DOI: 10.3390/math11112484

[28] Huang JX, Lee KS, Kim YK. Hybrid Translation with Classification: Revisiting Rule-Based and Neural Machine Translation. Electronics. 2020;9(2). DOI: 10.3390/electronics9020201