ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 2 - 2026

Research on Intelligent Translation and Cultural Adaptability of Cross-regional Earthquake Education Texts under the Framework of Federated Learning

Jing Chen[1,*]

[1] TCSOL Major, Tieling Normal College, Liaoning Province, 112008, China

SUMMARY: *This paper proposes a federated learning driven intelligent translation framework to solve the problems of corpus dispersion, non-uniform terminology and insufficient cultural adaptation in cross-regional earthquake education text translation. Based on Transformer, the framework integrates region Adapter, seismic term alignment and cultural adaptive feedback mechanism to realize multi-region collaborative training without uploading the original corpus. The experiment is based on 30600 bilingual sentence pairs for earthquake education. The results show that the BLEU of FedCA-NMT reaches 42.7, the COMET reaches 0.842, the term accuracy is 95.1%, and the cultural adaptability score is 87.4, which are better than those of the comparison models. Research shows that this method can improve the translation accuracy and action guidance effect of campus evacuation, community risk avoidance and disaster prevention for foreign residents.*

KEYWORDS: *Federated learning; Earthquake education text; Intelligent translation; Cultural adaptability*

1 Introduction

1.1 Research background and significance

Earthquake education texts are the basic information carriers in disaster risk communication, which mainly include pre-earthquake preparation, risk avoidance behavior, evacuation routes, emergency supplies, building safety, campus drills and community mutual assistance. This kind of text usually needs to be disseminated to audiences in different regions, different language backgrounds and different cultural experiences, and its translation quality directly affects the public's understanding, memory and action transformation of disaster knowledge. In the cross-regional communication scenario, the same earthquake education concept is often affected by regional experience, residential structure, school management system, community organization mode and audience language habits. For example, expressions such as "drop, cover and hold on", "temporary shelter" and "emergency kit" do not always have the same understanding effect through direct translation in different regions. If the localization context is ignored, the translation may have the problem of accurate terminology but unclear action direction.

Neural machine translation provides technical support for the cross-language transmission of disaster education texts. Multilingual pre-trained models and large-scale translation models have shown strong capabilities in general text translation. However, earthquake education

*xiaojing _lucky@163.com
https://doi.org/10.65102/is2026964 Published: 30 April, 2026

texts have the characteristics of dense field terms, clear command tone, sensitive risk semantics and strong cultural dependence. General translation models are prone to semantic weakening or cultural mismatch in texts from low-resource areas, localized expressions and emergency behavior descriptions. At the same time, cross-regional earthquake education corpus is often scattered in schools, emergency management departments, community platforms and public welfare organizations, and directly aggregating the original text may involve data rights, regional management norms and privacy protection issues. Federated learning can complete model co-training without centralizing original data, which provides a new technical path for cross-regional disaster education text translation [1-5]. Therefore, the combination of federated learning, neural machine translation, term alignment and cultural adaptability evaluation is helpful to improve the accuracy, safety and acceptability of earthquake education texts in cross-regional communication.

1.2 Research Objectives

Aiming at the problems of corpus dispersion, terminology inconsistency, insufficient cultural adaptation and limited data sharing in cross-regional earthquake education text translation, this study constructs a federated learning driven intelligent translation and cultural adaptation analysis framework. The research objectives include: (1) To establish a method for organizing cross-regional earthquake education text corpus, and to structurally label text topics, audience types, regional scenarios, risk behaviors and cultural expressions to form domain corpus resources suitable for model training; (2) A federated neural machine translation model for earthquake education texts is designed, so that nodes in different regions can share model parameters on the basis of retaining local data, so as to improve the semantic generalization ability of multi-regions. (3) A cross-lingual seismic term alignment method is constructed to reduce the deviation of professional terms, hazard avoidance actions and emergency facility names in translation; (4) A cultural adaptability evaluation mechanism is proposed to evaluate the quality of translation from the aspects of semantic accuracy, behavioral executability, regional expression fit and audience understanding.

This study focuses on three questions: RQ1, whether federated learning can improve the performance of cross-regional earthquake education text translation without aggregating the original corpus; RQ2, Can term alignment and cross-lingual semantic representation reduce the mistranslation rate of key concepts in disaster education texts? RQ3, Whether cultural adaptability evaluation can identify regionalized expression differences that are difficult to reflect by traditional translation indicators. Based on the above problems, the research hypothesis is as follows: H1, the federated trained model outperforms the single-region local model in terms of BLEU, COMET, and term accuracy. H2, after the introduction of the seismic term alignment module, the translation consistency of the expression of hazard avoidance actions, facility names and risk warning classes is significantly improved. H3, after adding the cultural adaptability score, the model has stronger understandability and action directivity of the translation in the text scenes oriented to schools, communities and foreign residents.

1.3 Research Framework

This research framework consists of five parts: corpus construction, federated translation modeling, semantic term alignment, cultural adaptation evaluation, and experimental validation. In the corpus construction part, data are collected from earthquake popular science materials, campus safety guidelines, community emergency manuals and multilingual disaster prevention propaganda texts, and annotated according to region, text function, audience object and cultural scene to form a cross-regional earthquake education text corpus. The federated translation modeling part is based on the Transformer encoder-decoder structure. The local training is completed in the client of each region, and then the model parameters are aggregated by the central server, so that the model can absorb the text features of different regions and avoid the cross-domain flow of the original corpus. The semantic term alignment part combines bilingual vocabulary, context embedding and domain term constraints to perform cross-language mapping of key expressions such as earthquake intensity, hazard avoidance action, evacuation space and emergency supplies, so as to enhance the professional stability of the translation.

The cultural adaptability evaluation part no longer only measures the quality of the translation by word similarity, but puts the translation into the specific disaster education situation to analyze whether the risk warning is clear, whether the behavioral steps are executable, whether the regional expression is natural, and whether there are obstacles for the audience to understand. In the experimental part, the centralized translation model, the local single-node model, the ordinary federated translation model and the model with term alignment and cultural adaptability constraints are set as the comparison objects, and the comprehensive evaluation is carried out by indicators such as BLEU, ROUGE, COMET, term accuracy, cultural adaptability score, communication overhead and training time. Through this framework, the research attempts to establish a more stable technical path between data security, translation performance and cultural adaptation, and provide methodological support for the intelligent dissemination of cross-regional earthquake education texts.

2 Literature Review

The existing research on the intelligent translation and cultural adaptability of cross-regional earthquake education texts can be generally divided into three categories. One is the research on the translation of earthquake education texts and the dissemination of disaster information, which focuses on how disaster knowledge is understood by audiences in different languages and regions. One is neural machine translation and cross-lingual text processing, which mainly improves the accuracy of translation through multilingual pre-training models, semantic representation and translation quality evaluation. The other is the application research of federated learning in natural language processing, which emphasizes the completion of model co-training under the condition of non-centralized flow of original data. The three categories of studies responded to disaster propagation, machine translation, and data security issues, respectively, but in the specific scenario of earthquake educational texts, the coordination between terminology unification, regional expression differences, cultural acceptability, and cross-node training efficiency still needs to be further addressed.

2.1 Research on translation of earthquake education texts and cross-regional disaster information dissemination

There is an obvious difference between the translation of earthquake education text and general information translation. Its core task is not only to convey the meaning of words, but also to ensure that the audience can accurately understand the risk avoidance behavior and risk instructions. Earthquake education materials usually include pre-earthquake preparation, building risk avoidance, campus evacuation, community meeting point, emergency kit configuration and post-earthquake self-rescue, etc. The texts are mandatory, informative and contextual. Petraroli and Baars[1] pointed out from the perspective of disaster preparedness and communication of foreign residents that the understanding effect of disaster information would be affected by language ability, life experience and social and cultural background. Koike et al. [2] conducted a survey on the large-scale earthquake preparedness of international students in Japan, which also showed that there were obvious differences in the acceptance of disaster knowledge among cross-cultural people. It can be seen that the translation of cross-regional earthquake education texts should not stay at the level of words and sentences, but also consider the audience's familiarity with risk avoidance scenes, public institutions and community organization methods.

The research on disaster education communication also emphasizes that the translation should have a clear action orientation. The study of İşleyen and Demirkaya[3] on disaster preparedness of post-earthquake nursing students showed that disaster response self-efficacy was closely related to knowledge understanding, training experience and situational judgment. If the action steps in the translated text are vague, the public may not be able to translate the information into effective behavior. Mota-Santiago et al. [4] proposed in the study of earthquake relief demand estimation that the disaster situation and resource demand of different urban areas are not the same, which puts forward requirements for the regional adaptation of educational texts. Adhikari et al. [5] study on seismic risk classification of school buildings also shows that risk expression in school scenarios needs to be combined with building types, student groups and management rules. Existing research provides a communication and disaster management basis for earthquake education text translation, but most of the results are still based on manual analysis and questionnaire survey, and lack of attention to computational modeling, terminology constraints and automated cultural adaptation of cross-lingual texts.

2.2 Neural Machine Translation and Cross-lingual Text Processing

Neural machine translation has become an important technical foundation for cross-language text processing. Multilingual pre-trained models can learn the semantic relationship between different languages in a shared representation space, thereby improving the performance of low-resource languages and cross-regional text translation. The large-scale multilingual translation model proposed by Costo-Jussa et al. [6] emphasizes human-oriented and low-resource language coverage, which provides technical reference for the multilingual transmission of disaster education texts. Compared with traditional statistical translation, neural machine translation can better capture contextual semantics, but in earthquake education texts, the model may still translate technical terms into common words, or process sentences with clear instruction meanings into weaker suggestion expressions. This kind of bias can weaken the risk warning function of the text in the disaster propagation scenario.

Translation quality evaluation provides further support for the training and validation of intelligent translation systems. Kocmi and Federmann[7] pointed out that large language models can be used to evaluate translation quality. The quality estimation method proposed by Rei et al. [8] can judge the reliability of translation without reference translation. Falcao et al. [9] discussed the application value of COMET in low-resource machine translation evaluation. WMT24 related research further shows that machine translation quality evaluation has entered a new stage of large model participation, but the translation problem has not been completely solved [10, 11]. These results show that it is difficult to fully evaluate the quality of earthquake education text translation by solely relying on BLEU, ROUGE and other word surface indicators. For expressions such as "avoid glass Windows", "enter the open area nearest to the school" and "evacuate according to the evacuation route of the school", the evaluation indicators should also pay attention to semantic accuracy, terminology consistency, behavior executable and regional cultural fit.

Cross-cultural acculturation research has received more attention in recent years. Singh et al. [12] discussed the role of language models in cultural adaptation, Cadotte et al. [13] focused on the value of cultural texts in low-resource language models, and Jinnai[14] pointed out that cross-cultural alignment may change the commonsense judgments of models. The dialect and language variety benchmark proposed by Faisal et al. [15] also suggests that regional language differences will affect the actual performance of natural language processing models. The above studies provide important implications for this paper: earthquake education text translation cannot only pursue general semantic similarity, but also need to identify regional expression habits, disaster experience differences, and audience understanding paths. The existing neural machine translation research is relatively mature in terms of general translation ability, but there is still room for expansion of term alignment, regional feature modeling and cultural adaptability evaluation in the field of earthquake education.

2.3 Research on the Application of Federated Learning in Natural Language Processing

Federated learning provides a suitable way of data collaboration for cross-regional earthquake education text translation. Earthquake education texts in different areas are often scattered in schools, communities, emergency management agencies and public welfare organizations. The direct concentration of original corpus may be limited by data rights, management norms and privacy protection. Federated learning enables regional nodes to participate in model optimization through local training and parameter aggregation, while reducing the cross-domain flow of original text. Weller et al. [16] studied pre-trained models in multilingual federated learning and showed that the federation mechanism can be combined with multilingual representations. Passban et al. [17] applied federated learning to the training of hybrid domain translation model and proved that the translation performance can be improved through parameter cooperation between different data domains. Nagy et al. [18] reviewed the application of federated learning in natural language processing from the perspective of privacy protection, emphasizing that this method is suitable for processing text tasks with data sensitivity.

At the technical level, federated learning is not simply to average multiple models, but to deal with data heterogeneity, communication cost, model convergence and privacy protection. El Ouadrhiri and Abdelhadi[19] pointed out that differential privacy can be used in deep learning and federated learning scenarios, but privacy noise may affect model performance. Abreha et al. [20] review on federated learning in edge computing shows that client computing power, communication frequency, and model scale all affect system availability. For cross-regional earthquake education text translation, there are differences in text topics, audience objects, language styles and cultural expressions in different regions. Directly adopting a unified aggregation strategy may lead to the dilution of regional features. Therefore, this paper focuses on term constraints, regional feature labeling, and culturally adaptive feedback in the federated translation model, so that the model retains the local expression characteristics while maintaining the cross-region generalization ability.

Existing research shows that federated learning has advantages in privacy protection and multi-node collaboration, neural machine translation has basic ability in cross-language semantic modeling, and disaster propagation research suggests that translations must have clear action values. Table 1 summarizes the related research types, technical characteristics and limitations. It can be seen that there is still a lack of close integration between the three types of research: there is a lack of cross-regional corpus organization methods for federal training in the field of disaster education. In machine translation research, seismic terms, hazard avoidance actions and regional cultural expressions are rarely included in the unified evaluation. Federated natural language processing research has paid insufficient attention to the application scenarios of disaster education texts. Based on these shortcomings, this paper integrates federated learning, intelligent translation and cultural adaptability evaluation into a research framework to improve the accuracy, interpretability and practical communication effect of cross-regional earthquake education text translation.

Table 1: Related research types, technical characteristics and their implications for this paper

Research Category	Representative Studies	Main Focus	Technical or Methodological Features	Existing Limitations	Implications for This Study
Earthquake education and disaster communication	Petraroli and Baars; Koike et al.; İşleyen and Demirkaya	Disaster understanding and public preparedness	Questionnaire survey and effect analysis	Lack of automated translation constraints	Translations should highlight action guidance
Earthquake risk and scenario expression	Mota-Santiago et al.; Adhikari et al.	Earthquake scenarios and rescue needs	Scenario modeling and risk classification	Limited integration with multilingual texts	Translations should fit regional scenarios
Neural machine translation and quality evaluation	Costa-Jussà et al.; Kocmi and Federmann; Rei et al.	Multilingual translation and quality assessment	Pre-trained models and COMET	Difficulty in evaluating cultural adaptability	Add terminology and cultural indicators
Cross-cultural and language variety processing	Singh et al.; Cadotte et al.; Faisal et al.	Cultural adaptation and language varieties	Large language models and variety benchmarks	Insufficient application to disaster texts	Identify regional expression differences
Federated learning and natural language processing	Weller et al.; Passban et al.; Nagy et al.	Collaborative training and privacy protection	Local training and parameter aggregation	High heterogeneity and communication cost	Support a federated translation framework

3 Research Methods

3.1 Construction of cross-regional seismic education text corpus and regional feature annotation

The intelligent translation of cross-regional earthquake education texts cannot directly apply the general bilingual corpus training method. A large number of expressions in earthquake education texts are related to emergency behavior, public space, school management, community risk avoidance and local life experience. If there is no regional feature annotation, the model is easy to misjudge "semantic correct" as "communication effective". Therefore, this paper takes corpus construction as the basis of the method, and integrates cross-regional text sources, text functions, audience types and cultural scenes into the annotation system at the same time, so that the subsequent federated translation model can learn the expression differences of different regions. The corpus construction and annotation process of this paper is shown in Figure 1. The core idea is to transform the earthquake education texts scattered in different regions and different use scenarios into structured corpus resources that can be calculated, managed, and federated trained.

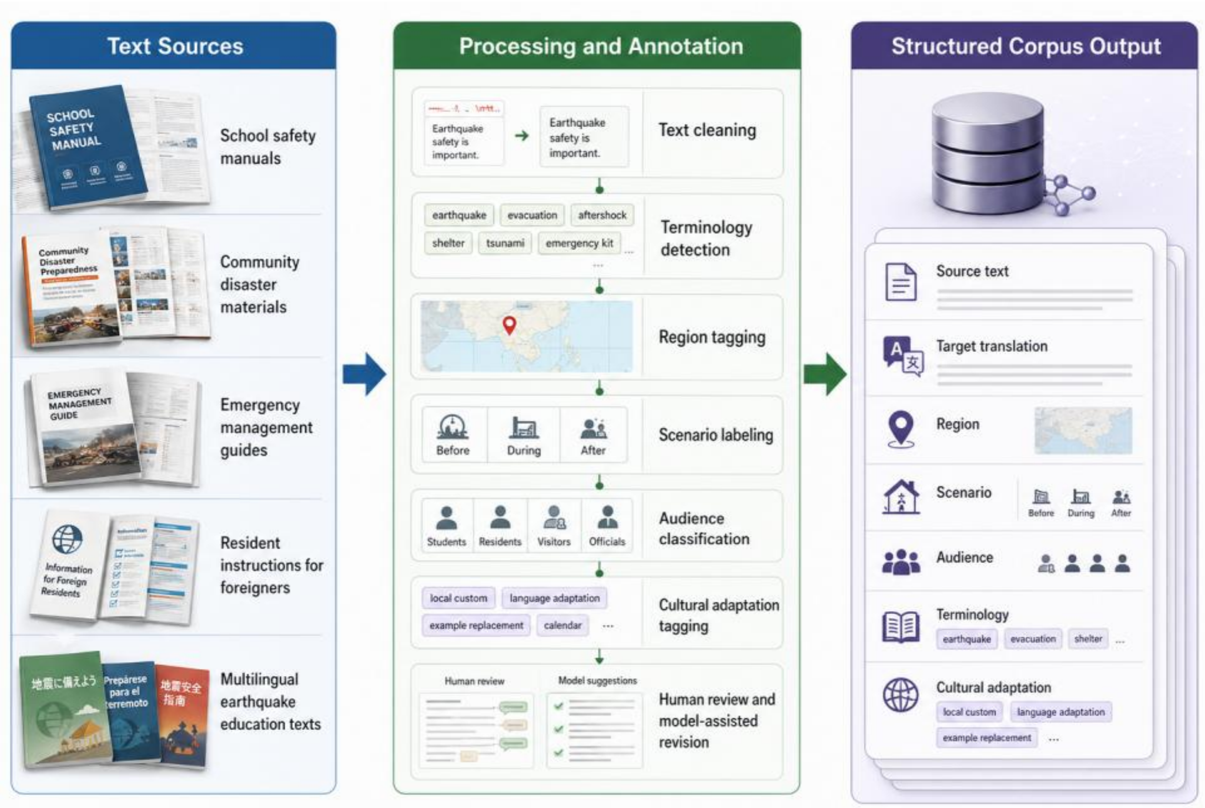

Figure 1: Process of cross-regional seismic education text corpus construction and regional feature annotation

The corpus collected in this paper includes five categories: school earthquake safety manual, community disaster prevention publicity materials, emergency management guide, disaster prevention instructions for foreign residents, and multilingual earthquake popular science texts. Each sample consists of source language text, target language translation, region label, scene label, term label and cultural adaptation label. Let the cross-regional earthquake education text corpus be:

$$D = \{(x_i, y_i, r_i, s_i, a_i, c_i)\}_{i=1}^{N} \tag{1}$$

where, x_i represents the source language text, y_i represents the target language translation, r_i represents the regional source, s_i represents the text usage scenario, a_i represents the audience type, and c_i represents the cultural adaptation annotation. Different from ordinary machine translation corpus, this structure not only preserves the bilingual text relationship, but also preserves the regional context attached to the text. The regional tags mainly record earthquake risk experience, building type, school organization mode and community shelter space. Scene tags include classroom education, campus drill, community announcement, resident manual and mobile tips. The audience labels include students, teachers, community residents, migrant workers and international residents.

In order to reduce the noise caused by the inconsistency of manual annotation, this paper adopts a three-stage process of "rule pre-labeling, manual verification and model-assisted correction". The pre-annotation part of the rules uses the earthquake term dictionary to identify terms such as "epicenter, intensity, aftershock, refuge, emergency kit, evacuation passage". In the manual verification part, annotators with disaster education or translation experience judged whether the translation conformed to the understanding habits of the target area audience. The model-aided correction part uses a pre-trained language model to identify potential low-consensus samples and returns them to the labeling side for review. The consistency of regional feature annotation was tested by Cohen's κ coefficient. If a batch was lower than 0.80, a double check was performed again. This process enables the corpus to have both a computable structure and retain the human judgment experience in earthquake education texts.

3.2 Architecture of intelligent translation model driven by Federated learning

Cross-regional earthquake education texts are usually saved by different regional institutions, and directly centralizing the corpus will increase the cost of data rights, privacy protection and management coordination. To solve this problem, this paper constructs an intelligent translation model architecture driven by federated learning. The model is based on the Transformer encoder-decoder, and the local training is completed at the client of each region, and the model parameters are aggregated by the central server. This method enables earthquake education texts from different regions to participate in model optimization, while the original corpus does not leave the local node. The overall model structure is shown in Figure 2, where the central server is responsible for global parameter maintenance and aggregation, and different regional clients are responsible for local corpus training and regional feature adaptation.

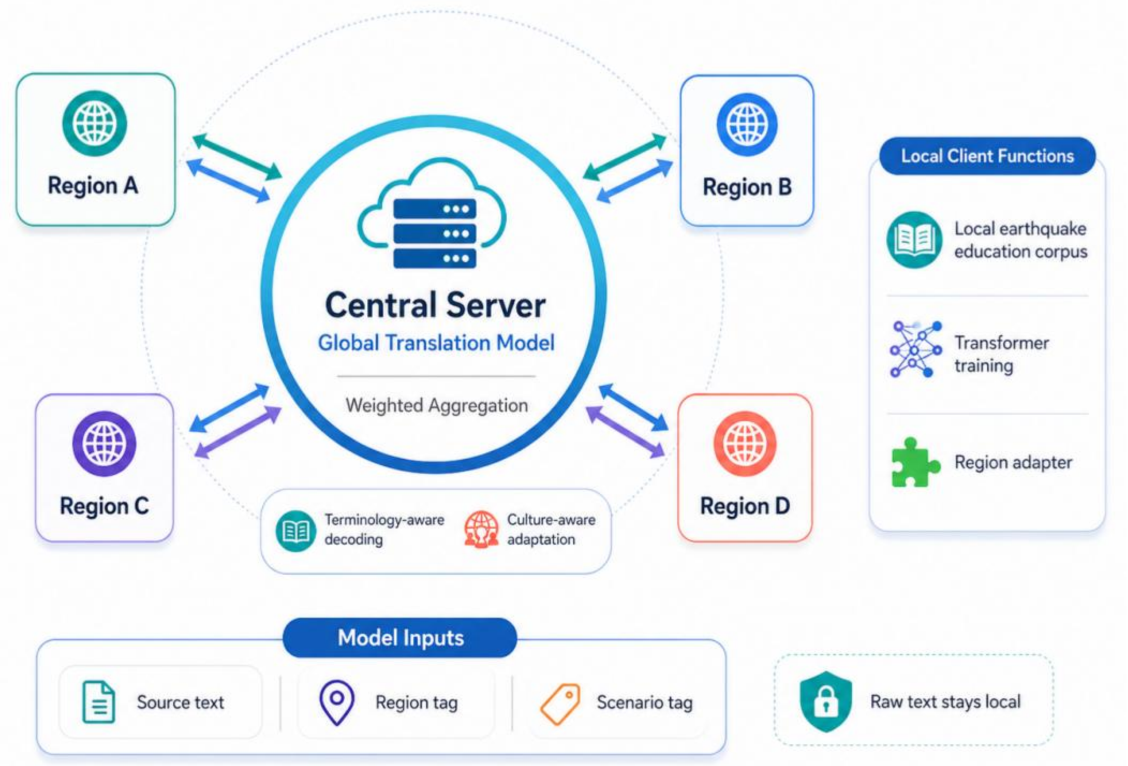

Figure 2: Architecture of intelligent translation model for seismic educational texts driven by federated learning

Let there be a total of K regional clients, the KTH client holds the local corpus D_k, and its sample size is n_k. The global model parameters are denoted as θ, and the federated training objective is denoted as follows.

$$\min_{\theta} F(\theta) = \sum_{k=1}^{K} \frac{n_k}{n} F_k(\theta), \quad n = \sum_{k=1}^{K} n_k \tag{2}$$

where $F_k(\theta)$ is the translation loss function at the KTH region node. In each round of training, the server sends the current global parameters to the selected regional clients, and each client completes several rounds of gradient updates with the local corpus, and then uploaders the model parameter difference or encrypted gradient. The server updates the global model using weighted aggregation:

$$\theta^{t+1} = \sum_{k=1}^{K} \frac{n_k}{n} \theta_k^{t+1} \tag{3}$$

This structure can alleviate the problem of cross-regional corpus imbalance. Nodes with larger sample size have higher weights in the global parameters, while regions with smaller sample size but special cultural expression retain local features through the regional adaptation layer. As shown in Figure 2, this paper adds a regional adaptation module to the base Transformer, each client configures lightweight Adapter parameters, the common layer learns the general semantics of earthquake education texts, and the regional layer learns local expressions and audience habits. In this way, the model does not have to train a complete translation system for each region, and does not force all regional expressions to be condensed into the same language style.

The input of the model consists of source text embedding, location encoding, region label embedding and scene label embedding. The encoder is responsible for extracting cross-lingual semantic features, the decoder generates the target language translation, and the term constraint module performs probabilistic correction of key seismic terms in the decoding stage. In the training process, cross entropy loss is used for common sentences, term constraint loss is added for termdense sentences, and region adaptation loss is added for cultural adaptation samples. The overall loss function is written as follows.

$$L = L_{mt} + \lambda_1 L_{term} + \lambda_2 L_{region} \tag{4}$$

where, L_{mt} represents machine translation loss, L_{term} represents term consistency loss, L_{region} represents region adaptation loss, and λ_1 and λ_2 are weight coefficients. Through this joint training method, the model can simultaneously focus on translation smoothness, term accuracy and region expression naturalness.

3.3 Cross-lingual semantic representation and seismic term alignment methods

The translation difficulties of earthquake education texts are not only at the sentence level, but also focused on terms, action phrases and risk warning expressions. The same concept may have multiple translations in different languages. For example, "shelter", "evacuation site" or "temporary refuge area" can be translated into "shelter", "evacuation site" or "temporary refuge area", but the corresponding institutional scenarios and action meanings of the three are not exactly the same. If the model generates the translation only based on the context probability, it is easy to produce expressions with similar word meanings but deviated action directions. Therefore, this paper introduces cross-lingual semantic representation and seismic term alignment method in the translation model. The main processing link of the proposed method is shown in Figure 3, which includes the links of term identification, context encoding, bilingual term base matching, semantic alignment and candidate ranking.

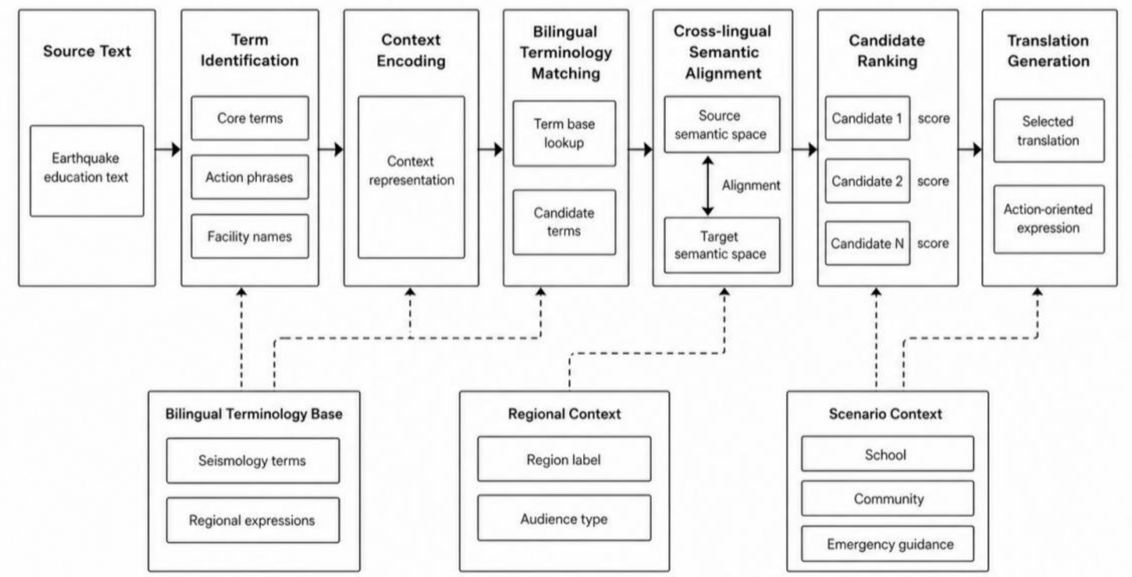

Figure 3: Cross-lingual semantic representation and seismic term alignment process

This paper constructs a two-layer term base. The base layer includes five categories: seismological terms, risk avoidance actions, building space, emergency materials and organization management. The regional layer records common expressions, names of public facilities and explanatory expressions for migrant residents in different regions. For each term t, the system records its source language form, target language candidate translation, semantic category, risk level, and applicable scenario. Term alignment does not use simple dictionary replacement, but combines context embedding to calculate the semantic distance between the source term and the target candidate expression. Let the source language term be denoted as $e_s(t)$ and the target language candidate be denoted as $e_t(z)$, and the alignment score between them be:

$$A(t,z) = \frac{e_s(t) \cdot e_t(z)}{\|e_s(t)\| \|e_t(z)\|} + \gamma R(t,z) \tag{5}$$

where, the first half is the cross-lingual semantic similarity, R(t,z) represents the term rule matching score, and γ is the rule constraint strength. If a candidate translation is close to the source term in the semantic space and satisfies the term category and scene constraints, it will obtain a higher alignment score. For action phrases such as "drop, cover and hold on", the system does not divide them into isolated words, but treats them as a complete risk-averse behavior unit to avoid confusion of action sequence or weakening of tone in the translation.

In the cross-lingual representation part, the multilingual pre-trained model is used to obtain context vectors, and then the domain continuing training is used to enhance the semantics of earthquake education. In this paper, the context of the sentence in which the term is located, the term category and the regional scene are jointly input into the representation layer, so that the model can distinguish between ordinary life words and earthquake education words. For example, "assembly point" can be understood as a meeting place in general context, but it is closer to "emergency meeting point" in earthquake education texts. After the domain semantic modification, the model can retain its emergency functional attributes when generating translations.

The term constraints are also embedded in the decoding stage. When the target sentence is generated to the position corresponding to the source term, the candidate word probability is affected by the term alignment score. The corrected generation probability is as follows.

$$P'(z) = \eta P(z|x) + (1 - \eta)A(t, z) \tag{6}$$

where P(z|x) is the original generation probability of the translation model, A(t,z) is the term alignment score, and η is the balance coefficient. This design avoids the blunt translation caused by pure rule substitution and also reduces the term drift that may occur in pure neural models.

3.4 Cultural adaptability evaluation and federated model training strategy

The quality of translation of earthquake education texts cannot be determined only by word surface similarity. Some translations score high on BLEU or ROUGE, but the target region audience may still have difficulty understanding the action requirements. For example, residents in some areas are not familiar with the evacuation system of specific schools, and the direct translation of "go to the designated assembly area" may lack necessary instructions; When facing foreign residents, the translation also needs to explain the expressions with local governance color such as "community", "grid worker" and "temporary settlement point". Therefore, this paper proposes a cultural adaptability evaluation mechanism to judge the actual communication effect of the translation by placing it in the disaster education scenario.

The evaluation of cultural adaptability consists of four dimensions: semantic accuracy, action executability, regional expression fit and audience understanding friendliness. Semantic accuracy measures whether the translation completely retains the disaster knowledge in the source text. Action executability measures whether the translation can clearly guide the audience to complete the risk avoidance action. Regional expression fit measures whether the translation conforms to the public space, school system and community organization mode of the target area. Audience understanding friendliness measures whether the translation avoids over-specialized expression. Let the comprehensive score of cultural adaptability be:

$$C = \alpha C_s + \beta C_a + \delta C_r + \rho C_u \tag{7}$$

where C_s, C_a, C_r, C_u represent the four scores of semantic accuracy, action executability, regional fit and understanding friendliness respectively, and α, β, δ and ρ are the weight parameters. Each weight was calibrated by the consistency of manual scoring on the validation set, so that the evaluation results were closer to the real use requirements of disaster education texts. As shown in Figure 4, after the four evaluation results are summarized by the cultural feedback module, they will act in reverse on the federated model update stage, so as to further promote "translation right" to "translation usable and easy to understand".

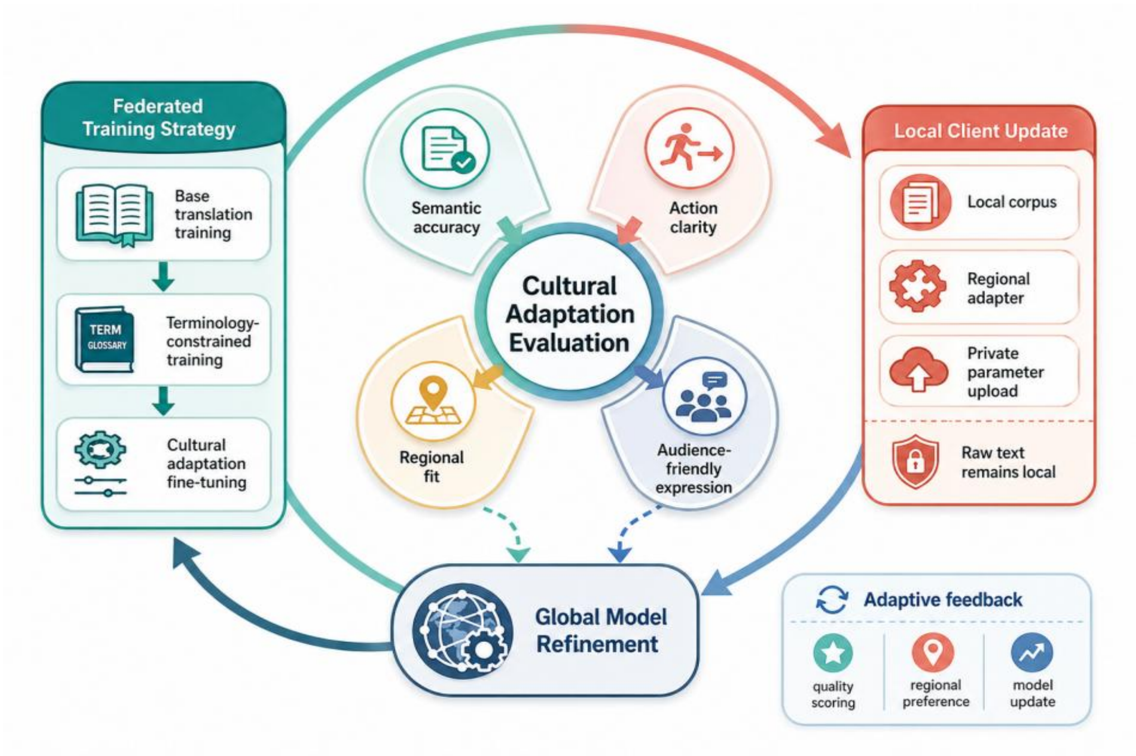

Figure 4: Cultural adaptability evaluation and federated training strategy

In the federated training stage, we adopt a segmentation strategy of "basic translation training-term constraint training-cultural adaptation fine-tuning". Basic translation training is responsible for building cross-language transfer ability; Terminology constraint training strengthens seismic professional expression and avoiding action phrases; Cultural adaptation fine-tuning uses samples with regional labels and human ratings to optimize the translation style. To reduce the communication overhead, the client does not upload the parameters in each batch, but completes several rounds of updates locally before participating in the aggregation. For nodes with small sample size but obvious regional characteristics, the system increases their cultural adaptation loss weight to avoid their expression being covered by large sample areas.

In terms of privacy protection, the uploaded content of the model only includes gradient or parameter updates, and the original text is not uploaded. Differential privacy noise can be added to sensitive nodes, and the parameter update form is as follows.

$$\Delta\tilde{\theta}_k = \Delta\theta_k + \mathcal{N}(0, \sigma^2) \tag{8}$$

where, $\Delta\tilde{\theta}_k$ represents the local model update after adding the privacy perturbation, $\Delta\theta_k$ represents the local model update, and $\mathcal{N}(0, \sigma^2)$ represents the Gaussian noise. In order to reduce the influence of noise on term translation, privacy perturbation is mainly applied to the general model layer, and a small perturbation range is set for the term constraint table and region adaptation parameters. This strategy takes into account data security and translation stability.

4 Experimental Setup

4.1 Dataset construction

In order to verify the effectiveness of the intelligent translation and cultural adaptation method of cross-regional earthquake education texts under the framework of federated learning, this paper constructs a bilingual corpus for cross-regional earthquake education. The corpus sources include school earthquake safety education manuals, community disaster prevention publicity materials, public guidelines for emergency management departments, disaster prevention instructions for foreign residents, and multilingual earthquake science texts. The text content covers the topics of pre-earthquake preparation, classroom risk avoidance, campus evacuation, community gathering, temporary settlement, emergency supplies, post-earthquake self-rescue and psychological comfort. In the data cleaning stage, repeated sentences, short sentences, advertising instructions and non-educational texts were deleted, and bilingual sentence pairs were manually verified.

In the experiment, the corpus was divided into five regional clients, corresponding to campus education node, community publicity node, emergency management node, foreign residents service node and multilingual science popularization node. Each node only holds the local text and does not upload the original corpus to the central server. The dataset is divided into training set, validation set and test set according to 7:1.5:1.5, and the distribution of different topics, audiences and text lengths is basically the same. Table 2 lists the construction of the dataset in this paper.

Table 2: Composition of cross-regional seismic education text dataset

Regional Client	Main Text Sources	Number of Sentence Pairs	Average Sentence Length / Words	Main Audience	Thematic Focus
C1 Campus Education Node	School safety manuals, drill notices	6,420	23.6	Students, teachers	Classroom emergency avoidance, campus evacuation
C2 Community Publicity Node	Community announcements, disaster prevention brochures	5,860	21.8	Community residents	Shelters, neighborhood mutual assistance
C3 Emergency Management Node	Emergency guidelines, policy explanations	7,240	28.4	Administrators, the public	Early-warning response, resettlement procedures
C4 Foreign Resident Node	Foreign resident manuals, multilingual instructions	4,780	19.7	Foreign residents	Basic terminology, daily-life scenarios
C5 Science Communication Node	Earthquake science webpages, publicity materials	6,300	25.1	General public	Earthquake knowledge, self-rescue methods

In order to ensure the credibility of the evaluation results, this paper selected an additional 1,000 groups of samples from the test set to construct a cultural adaptability test subset, including 300 groups of school scenarios, 250 groups of community scenarios, 220 groups of migrant residents scenarios, and 230 groups of emergency guidance scenarios. Each group of samples was scored by three reviewers with translation or disaster education background, and the scoring consistency was 0.79 using Fleiss 'κ test, indicating that the labeling results had high consistency.

4.2 Evaluation indicators

The experiment uses the combination of machine translation general indicators and domain adaptation indicators to evaluate the effect of the model. Generic metrics include BLEU, ROUGE-L, and COMET, which measure the degree of lexical coincidence, sequence similarity, and semantic agreement between the translation and the reference translation. Considering that earthquake education texts have strong term dependence, this paper further sets the term accuracy rate to investigate whether earthquake terms, avoidance actions and emergency facility expressions are correctly translated. Term accuracy is defined as:

$$\mathrm{TAR} = \frac{\mathrm{Ncorrect}}{\mathrm{Nterm}} \times 100 \tag{9}$$

where, $\mathrm{Ncorrect}$ denotes the number of correctly translated terms in the translation and Nterm denotes the total number of terms that should be translated in the test set. For high-frequency terms such as "aftershock", "refuge place", "evacuation passage", "emergency package" and "local risk avoidance", if the translation semantics is correct but the scene direction is not clear, it will not be counted as completely correct.

Cultural adaptability was evaluated using a four-dimensional score, including semantic accuracy, action executability, regional expression fit and audience understanding friendliness. The overall score is expressed as follows.

$$\mathrm{CAS} = 0.30\mathrm{C_s} + 0.30\mathrm{C_a} + 0.20\mathrm{C_r} + 0.20\mathrm{C_u} \tag{10}$$

Among them, $\mathrm{C_s}$ represents semantic accuracy, $\mathrm{C_a}$ represents action executability, $\mathrm{C_r}$ represents region expression fit, and $\mathrm{C_u}$ represents audience understanding friendliness. Each dimension is scored on a scale of 0-100. Since federated learning involves multi-node training, this paper also records single round communication volume, total communication rounds and training convergence time, which are used to evaluate the computational cost of the model in practical deployment. The communication cost is defined as follows.

$$\mathrm{Cost} = \sum_{t=1}^{T} \sum_{k=1}^{K} S(\Delta\theta_k^t) \tag{11}$$

where, T represents the communication round, K represents the number of clients, and $S(\Delta\theta_k^t)$ represents the amount of data uploaded by the KTH client for parameter updates in the TTH round.

4.3 Comparing Models

In order to analyze the effectiveness of the proposed method, a variety of comparison models are set up, including the traditional statistical translation method, the centralized neural machine translation model, the local single-node model, the ordinary federated translation model and the proposed method. Table 3 shows the main configurations of each model. All models use the same training set, validation set and test set, and keep the same word segmentation method, maximum sentence length and evaluation script to reduce the influence of non-model factors on the results.

Table 3: Comparison of models and experimental configurations

Model Name	Training Mode	Core Method	Term Constraints Used	Cultural Adaptation Feedback Used	Main Purpose
Dictionary-MT	Non-neural method	Bilingual dictionary matching and rule replacement	Yes	No	Traditional baseline
Transformer-Central	Centralized training	Standard Transformer translation model	No	No	Centralized training comparison
Local-Transformer	Single-node training	Independently trained Transformer for each region	No	No	Local model comparison
FedAvg-NMT	Federated training	FedAvg aggregation and Transformer	No	No	Common federated comparison
FedTerm-NMT	Federated training	Federated translation with term constraints	Yes	No	Term-enhanced comparison
FedCA-NMT	Federated training	Federated translation with term alignment and cultural adaptation feedback	Yes	Yes	Proposed method

Dictionary-MT is mainly used to test the upper limit and deficiency of rule translation in disaster education texts. Transformer-Central is used to observe performance under centralized training conditions. Local-Transformer is used to analyze the generalization ability of regional corpora when trained independently. FedAvg-NMT tests whether ordinary federated learning can improve the collaborative effect of multi-regions. FedTerm-NMT is used to verify the contribution of term constraints individually. FedCA-NMT, which simultaneously adds seismic term alignment and cultural adaptive feedback on the basis of federated training, is the final model of this paper.

4.4 Implementation Details

The experimental environment uses Ubuntu 22.04 operating system, Python 3.11, PyTorch 2.2 and Transformers 4.38. The hardware configuration is Intel Core i7-12700 processor, 32 GB memory, and NVIDIA RTX 4080 16 GB GPU. The model adopts a 6-layer encoder and 6-layer decoder structure, the hidden layer dimension is 768, the number of attention heads is

12, the feedforward layer dimension is 3072, and the dropout is set to 0.1. The maximum input length is set to 128, the batch size is 32, the optimizer is AdamW, the initial learning rate is 2×10-4, and the weight decay is 0.01.

Five clients are set for federated training, 80% of them are randomly selected to participate in aggregation in each round, the number of local training rounds is set to 2, and the number of global communication rounds is set to 80. The term constraint loss weight λ1 is set to 0.4, and the region adaptation loss weight λ2 is set to 0.3. In order to avoid small-scale regional nodes being covered by large-scale nodes, a regional equilibrium coefficient is added to the sample weight during aggregation, so that foreign resident nodes and multilingual science popularization nodes maintain sufficient influence in the cultural adaptability test. The validation set COMET triggers early stop when there is no promotion for 5 consecutive rounds. All experiments were repeated three times, and the results were averaged to reduce the impact caused by random initialization and differences in client sampling.

5 Results and analysis

5.1 Cross-region translation performance comparison

The experimental results show that the federated learning framework can effectively improve the translation quality of cross-regional earthquake education texts, especially in terms of terminology accuracy and cultural adaptability. Table 4 presents the overall comparison results of different models on the test set. Compared with Dictionary-MT, Local-Transformer, and plain FeDagu-NMT, FedCA-NMT achieves the highest results in BLEU, ROUGE-L, COMET, term accuracy, and cultural adaptability scores. Among them, the BLEU of FedCA-NMT reaches 42.7, which is 4.5 higher than that of FEDagu-NMT. COMET reaches 0.842, which is 0.081 higher than that of centralized Transformer. The term accuracy reaches 95.1%, indicating that the seismic term alignment module can effectively reduce the mistranagram of high-risk expressions such as "refuge place", "aftershock" and "evacuation passage".

Table 4: Performance comparison of different models on the cross-regional seismic education text test set

Model	BLEU	ROUGE-L	COMET	Terminology Accuracy / %	Cultural Adaptability Score
Dictionary-MT	28.6	44.2	0.641	78.4	62.1
Transformer-Central	36.9	54.8	0.761	84.7	73.5
Local-Transformer	33.4	50.2	0.724	82.1	76.8
FedAvg-NMT	38.2	56.1	0.787	86.5	77.9
FedTerm-NMT	40.5	58.0	0.811	93.2	80.6
FedCA-NMT	42.7	60.4	0.842	95.1	87.4

Taking FEDARG-NMT as the benchmark, the improvement rate of FedCA-NMT in cultural adaptability score is 12.19%, which is significantly higher than that of BLEU, indicating that ordinary translation indicators are difficult to fully reflect the regional dissemination effect of earthquake education texts. The results of paired t-test showed that FedCA-NMT had statistically significant improvements in COMET, term accuracy and cultural adaptability scores compared with FeDARG-NMT, with p values less than 0.01.

From the perspective of nodes in different regions, the improvement of the model on foreign residents nodes and community publicity nodes is more prominent. Figure 5 illustrates the BLEU and cultural adaptability scores of FEDAV-NMT compared with FedCA-NMT on the five regional clients. It can be seen that the BLEU of C4 foreign resident node is improved from 35.4 to 40.6, and the cultural adaptability score is improved from 72.3 to 86.1, indicating that the regional adaptation module can alleviate the problem of insufficient explanation of life scenes in the texts of foreign residents.

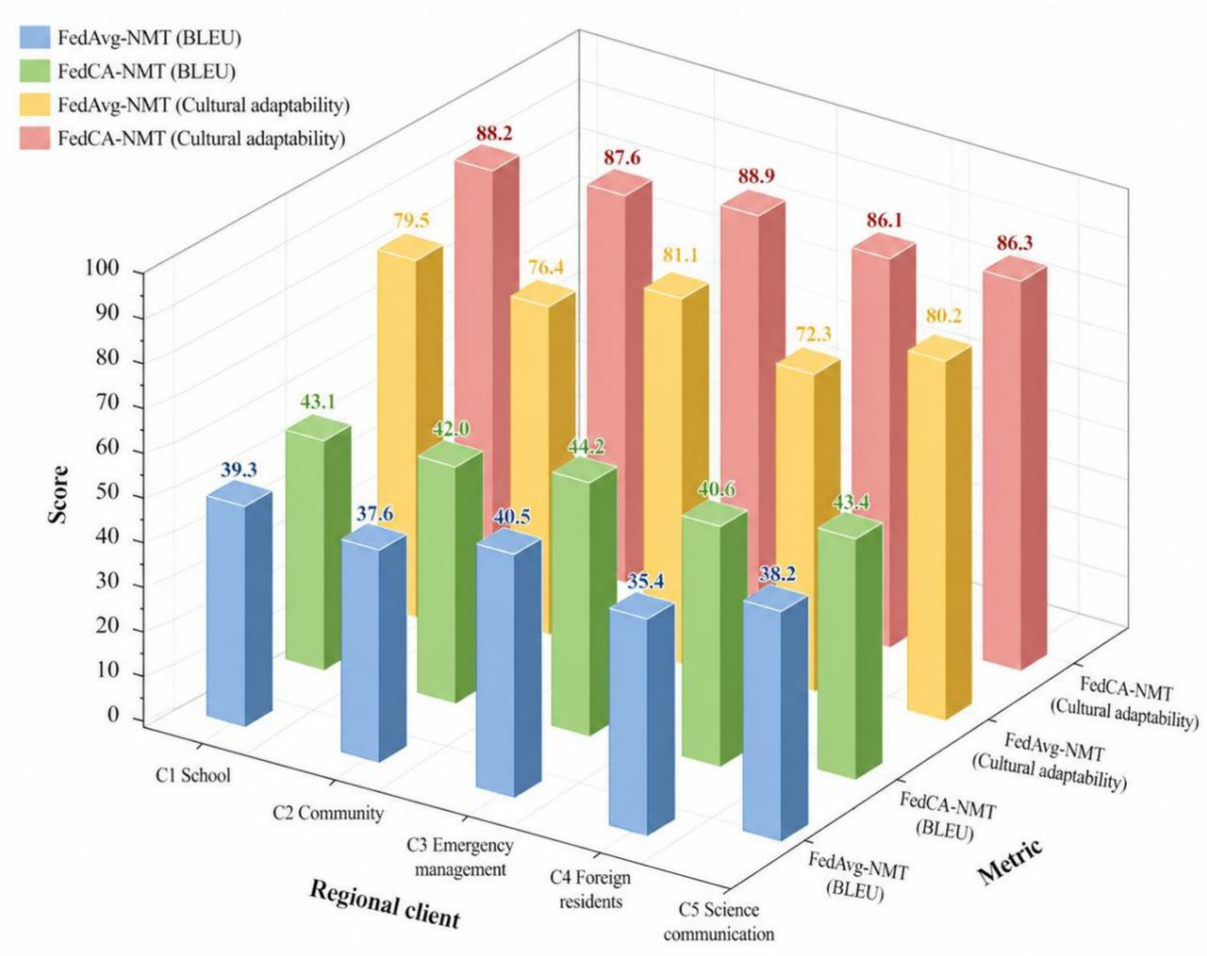

Figure 5: Comparison of client translation performance in each region

5.2 Analysis of ablation experiments

In order to further judge the contribution of each module to the model performance, this paper conducts ablation experiments based on FedCA-NMT. The ablation objects include federated aggregation mechanism, term alignment module, region Adapter, culturally adaptive feedback, and region balanced aggregation weights. The experimental results are shown in Table 5. After removing the cultural adaptability feedback, the BLEU only decreases by 1.4, but the cultural adaptability score decreases by 7.8, indicating that the module mainly affects the communication usability of the translation, rather than the simple word similarity. After removing the term alignment module, the term accuracy decreases from 95.1% to 86.7%, which is the group with the most significant decline in term performance among all ablation configurations.

Table 5: Results of FedCA-NMT ablation experiments

Configuration	BLEU	COMET	Terminology Accuracy / %	Cultural Adaptability Score	CAS Decrease
Full FedCA-NMT	42.7	0.842	95.1	87.4	—
Without federated aggregation	39.1	0.794	92.0	80.2	-7.2
Without terminology alignment	39.8	0.807	86.7	82.1	-5.3
Without regional Adapter	40.6	0.819	94.2	82.9	-4.5
Without cultural adaptation feedback	41.3	0.825	94.8	79.6	-7.8
Fixed aggregation weights	41.0	0.818	94.5	84.2	-3.2

From the perspective of module contribution, term alignment module mainly improves professional stability, region Adapter mainly improves expression adaptation of different clients, and cultural adaptation feedback improves action hints and audience understanding quality. Figure 6 further presents the cultural adaptation changes of different ablation configurations in three types of typical text scenarios. The school scene is more sensitive to the sequence of actions, the community scene is more sensitive to spatial names and organizational relationships, and the foreign resident scene requires more explanatory expression. After removing the acculturation feedback, the score of the texts of foreign residents decreased most significantly, from 86.1 to 76.8, indicating that the cross-cultural audience relied more on the background supplement and expression transformation in the translation.

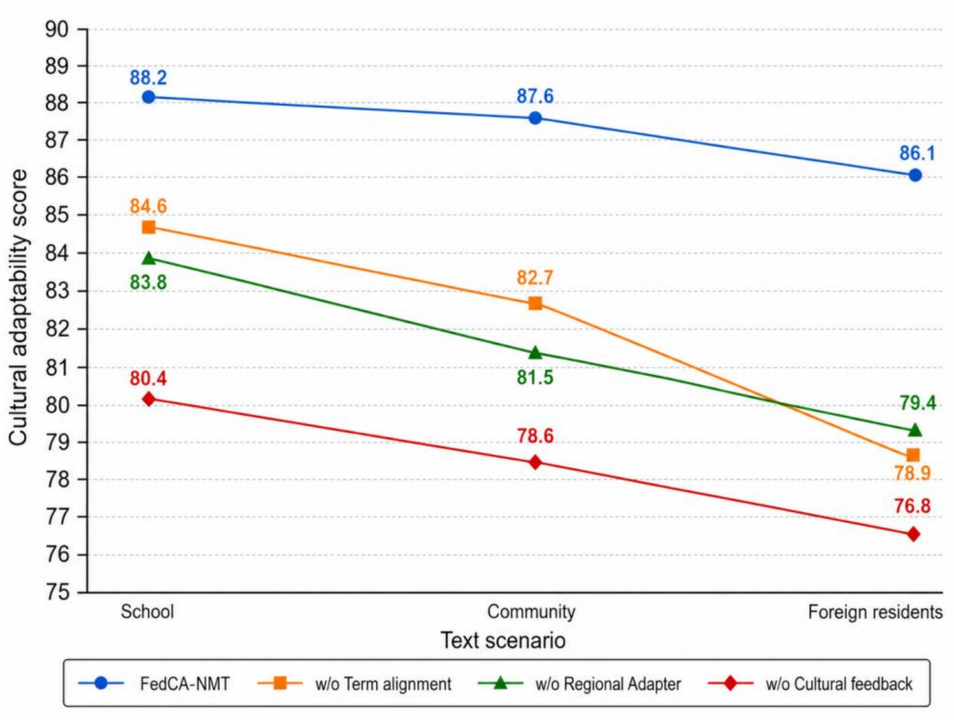

Figure 6: Cultural adaptability scores of different ablation configurations in typical scenarios

5.3 Analysis of computational efficiency and communication overhead

While improving the performance of federated translation models, training efficiency and communication cost should also be considered. The experiments record the validation set COMET changes and cumulative communication overhead of FeDAV-NMT, FedTerm-NMT, and FedCA-NMT over 80 rounds of federated training. Figure 7 illustrates the convergence trends of the three federated models under different communication rounds. It can be seen that the COMET of FedCA-NMT has reached 0.826 at the 40th round, which is higher than 0.801 of FedTerm-NMT and 0.776 of FeDARG-NMT. In the 60th round, FedCA-NMT was further improved to 0.841, and stabilized at 0.842 in the 80th round, indicating that term alignment and acculturation feedback can accelerate the model to enter a stable state. In contrast, the COMET of FeDARG-NMT at round 80 is only 0.787, and that of FedTerm-NMT is 0.811, both of which are lower than FedCA-NMT.

Figure 7 simultaneously presents the cumulative communication overhead of different models. At the end of the 80th training round, the cumulative communication volume of FeDAV-NMT is 9.44 GB, that of FedTerm-NMT is 10.08 GB, and that of FedCA-NMT is 10.72 GB. FedCA-NMT has higher cumulative communication traffic than FEDagu-NMT due to the addition of regional Adapter update and acculturation feedback parameters, but its COMET improvement is more obvious. From the average traffic volume of a single round, FedCA-NMT is about 134 MB, FedTerm-NMT is about 126 MB, and FeDARG-NMT is about 118 MB. If the full model parameters are uploaded, the communication volume of a single round is about 421 MB. After using the shared layer gradient and region Adapter update, the single-round communication volume of FedCA-NMT is reduced to 134 MB, and the communication overhead is reduced by about 68.17%.

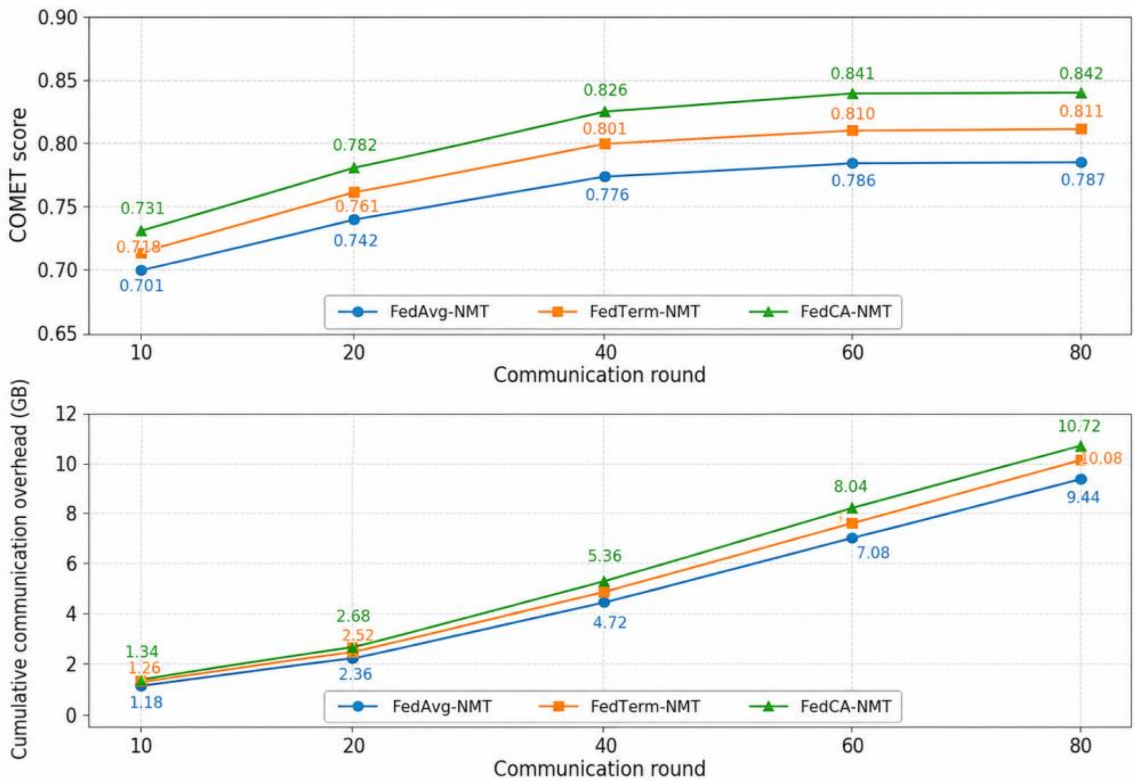

Figure 7: Convergence trend and cumulative communication overhead of different federated models

In terms of training time, the total training time of FeDAV-NMT is 11.6 h, FedTerm-NMT is 12.4 h, and FedCA-NMT is 13.7 h. Compared with FeDagu-NMT, the training time of FedCA-NMT was increased by 18.1%, but the cultural adaptability score was increased from 77.9 to 87.4, with an improvement rate of 12.19%. The term accuracy is improved from 86.5% to 95.1%, an increase of 8.6 percentage points. It can be seen that although FedCA-NMT brings some additional communication and training costs, it can obtain higher semantic consistency, term stability and regional cultural adaptation ability, which is suitable for cross-regional collaborative translation of disaster education texts.

6 Discussion

The experimental results show that the federated learning framework can improve the intelligent translation quality of earthquake education texts without centralizing the cross-regional original corpus. Compared with the centralized Transformer and the ordinary FeDAV-NMT, the advantages of FedCA-NMT are not only reflected in the general indicators such as BLEU and COMET, but also reflected in the term accuracy and cultural adaptability score. The results show that the translation of earthquake education text cannot only rely on general semantic modeling, but also need to incorporate the expression of risk avoidance actions, emergency facilities, school evacuation system and community governance into the model constraints. The term alignment module reduces the translation drift of core concepts such as "refuge place", "evacuation channel" and "emergency package", and the region Adapter enables the model to retain the expression habits of different regions, avoiding the problem of text style being averaged after cross-region aggregation.

From the perspective of methodological significance, the proposed model deals with data security, translation performance and propagation adaptation in the same framework. The federal training solves the problem that the corpus of schools, communities and emergency management agencies is difficult to be centralized and shared. The cultural adaptability evaluation makes up for the shortcomings of traditional machine translation indicators that lack of description of the executability of actions and the understanding of audiences. For texts such as disaster prevention instructions for foreign residents, campus drill notices and community emergency announcements, whether the translation is "understandable, executable, and corresponding to the local scene" is often more important than simple word similarity. Therefore, this paper introduces the acculturation feedback into the model training, which further extends the translation evaluation from the language level to the disaster education communication effect level.

At the same time, this paper still has some limitations. On the one hand, cultural adaptability scores still rely on manual review samples. Although the scoring consistency has reached a high level, more automated and interpretable evaluation mechanisms are still needed in larger corpora. On the other hand, FedCA-NMT adds term constraints and regional adaptation parameters, and the training time and communication overhead are higher than the ordinary federated model, which needs to further compress the model size when deployed in resource-constrained areas. Subsequent research can combine knowledge distillation, efficient parameter fine-tuning and dynamic communication compression methods to reduce the computational pressure of the client. At the same time, the model is extended to more languages, more earthquake experience areas and real disaster drill texts to verify the stability of the model in complex communication scenarios.

7 Conclusions

Focusing on the problems of data dispersion, term inconsistency and insufficient cultural adaptation in cross-regional earthquake education text translation, this paper constructs a federated learning driven intelligent translation framework. This study integrates seismic education corpus annotation, Transformer translation model, seismic term alignment, regional Adapter and cultural adaptability evaluation, so that different regional nodes can participate in model training without uploading the original text. Experimental results show that FedCA-NMT is superior to the comparison models in terms of BLEU, COMET, term accuracy and cultural adaptability scores, especially in the texts of campus evacuation, community risk avoidance and disaster prevention for foreign residents, FEDCA-NMT shows stronger action directivity and regional expression adaptation ability. This study shows that the intelligent translation of disaster education texts should not only pursue the equivalence at the language level, but also pay attention to whether the translation is easy to understand, whether it can guide actions, and whether it conforms to the public service context of the target area. Limited by the scale of corpus, the cost of manual evaluation and the computing power of the client, subsequent research still needs to expand more languages and real drill texts, and combine model compression and dynamic communication optimization to improve the efficiency of system deployment.

References

[1] Weller O, Marone M, Braverman V, et al. Pretrained models for multilingual federated learning[C]//Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2022: 1413-1421.

[2] Passban P, Roosta T, Gupta R, et al. Training mixed-domain translation models via federated learning[C]//Proceedings of the 2022 conference of the north american chapter of the association for computational linguistics: Human language technologies. 2022: 2576-2586.

[3] Nagy B, Hegedűs I, Sándor N, et al. Privacy-preserving federated learning and its application to natural language processing[J]. Knowledge-Based Systems, 2023, 268: 110475.

[4] El Ouadrhiri A, Abdelhadi A. Differential privacy for deep and federated learning: A survey[J]. IEEE access, 2022, 10: 22359-22380.

[5] Abreha H G, Hayajneh M, Serhani M A. Federated learning in edge computing: a systematic survey[J]. Sensors, 2022, 22(2): 450.

[6] Costa-Jussà M R, Cross J, Çelebi O, et al. No language left behind: Scaling human-centered machine translation[J]. arXiv preprint arXiv:2207.04672, 2022.

[7] Kocmi T, Federmann C. Large language models are state-of-the-art evaluators of translation quality[C]//Proceedings of the 24th Annual Conference of the European Association for Machine Translation. 2023: 193-203.

[8] Rei R, Treviso M, Guerreiro N M, et al. CometKiwi: IST-unbabel 2022 submission for the quality estimation shared task[C]//Proceedings of the Seventh Conference on Machine Translation (WMT). 2022: 634-645.

[9] Falcão J, Borg C, Aranberri N, et al. COMET for low-resource machine translation evaluation: A case study of English-Maltese and Spanish-Basque[C]//Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024). 2024: 3553-3565.

[10] Kocmi T, Avramidis E, Bawden R, et al. Findings of the WMT24 general machine translation shared task: The LLM era is here but MT is not solved yet[C]//Proceedings of the Ninth Conference on Machine Translation. 2024: 1-46.

[11] Freitag M, Mathur N, Deutsch D, et al. Are LLMs breaking MT metrics? results of the WMT24 metrics shared task[C]//Proceedings of the Ninth Conference on Machine Translation. 2024: 47-81.

[12] Singh P, Patidar M, Vig L. Translating across cultures: Llms for intralingual cultural adaptation[C]//Proceedings of the 28th Conference on Computational Natural Language Learning. 2024: 400-418.

[13] Cadotte A, André N, Sadat F. Machine Translation Through Cultural Texts: Can Verses and Prose Help Low-Resource Indigenous Models?[C]//Proceedings of the Seventh Workshop on Technologies for Machine Translation of Low-Resource Languages (LoResMT 2024). 2024: 121-127.

[14] Jinnai Y. Does cross-cultural alignment change the commonsense morality of language models?[C]//Proceedings of the 2nd Workshop on Cross-Cultural Considerations in NLP. 2024: 48-64.

[15] Faisal F, Ahia O, Srivastava A, et al. DIALECTBENCH: An NLP benchmark for dialects, varieties, and closely-related languages[C]//Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2024: 14412-14454.

[16] Petraroli I, Baars R C. Disaster preparedness communication and perception of foreign residents in Kansai, Japan: A socio-cultural study[J]. TheAustralian Journal of Emergency Management, 2022, 37(2): 81-89.

[17] İşleyen E K, Demirkaya Z. Relationship between disaster response self-efficacy and disaster preparedness in nursing students: after-earthquake study[J]. Disaster medicine and public health preparedness, 2024, 18: e83.

[18] Koike R, Sonoda N, Furuki H, et al. Individual preparedness for large-scale earthquakes among international students in Japan: a cross-sectional questionnaire survey[J]. JMA journal, 2024, 7(4): 496-505.

[19] Mota-Santiago L R, Lozano A, Ortiz-Valera A E. Determination of disaster scenarios for estimating relief demand to develop an early response to an earthquake disaster in urban areas of developing countries[J]. International Journal of Disaster Risk Reduction, 2023, 87: 103570.

[20] Adhikari R K, D'Ayala D, Fernandez R, et al. GLOSI taxonomy: a tool for ‘seismic risk assessment’oriented classification of school buildings[J]. International Journal of Disaster Risk Reduction, 2023, 87: 103594.