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

Optimization of Communication Content of Chinese Traditional Culture to the Outside World Based on Vector Algorithms

Qi Tao[1,*]

[1] International Cooperation Office, Hongkong, Macao and Taiwan Af'fairs Office, Donghua University, Shanghai, 201620, China

SUMMARY: *Aiming at the problems of content homogeneity, semantic deviation and insufficient cross-cultural adaptation in the external communication of traditional Chinese culture, this paper constructs an integrated research framework for communication content recognition, vector modeling and optimal output. This paper uses text embedding, semantic similarity calculation, topic clustering, cross-language semantic mapping and other methods to establish a multi-level representation system of word level, sentence level and topic level cooperation, and designs the mechanism of content screening, semantic matching, topic aggregation and expression optimization. The results show that the clustering compactness of the art texts of intangible cultural heritage reaches 0.88, and the inter-class separation of the classic thought texts reaches 0.84. The semantic consistency of the optimized text was improved from 0.74 to 0.87, the communication clarity was improved from 0.69 to 0.83, and the cross-cultural fitness was improved from 0.66 to 0.81. The semantic recognition accuracy, content recommendation accuracy and comprehensive performance score of the proposed method reach 0.89, 0.86 and 0.87, respectively, which are better than those of keyword matching, rule method and general text classification method. The research shows that the vector algorithm can effectively improve the semantic organization ability, expression stability and international communication adaptation level of traditional cultural communication content.*

KEYWORDS: *International dissemination of traditional culture; Vector algorithm; Semantic representation; Content optimization*

1 Introduction

Under the background of accelerating the reconstruction of the global communication system and the deep involvement of digital media in international cultural exchanges, the external communication of traditional Chinese culture has gradually shifted from the relatively single translation and export in the past to the compound communication form of multi-platform, multi-language and multi-scene coordination. Traditional culture not only carries the historical memory, values and aesthetic paradigms of Chinese civilization, but also is an important resource for national cultural image construction, cultural exchanges and mutual learning and international discourse expression. However, from the perspective of actual communication practice, there are still prominent structural problems in the organization and presentation of traditional cultural content in international communication: Some texts have been focused on high-frequency elements such as festival customs, instrument symbols, and

*guizhoumeixi@163.com
https://doi.org/10.65102/is2026973
Published: 30 April, 2026

classical imagery for a long time. The content selection tends to be concentrated, and the narrative mode is relatively solidified, which leads to high homogeneity in theme composition and meaning expression of different communication texts, and it is difficult to form a content system with clear layers and outstanding recognition. At the same time, the ethical spirit, historical context and symbolic meaning carried by traditional culture are significantly dependent on local culture. In the process of cross-language conversion and cross-cultural interpretation, semantic compression, connotation loss and interpretation deviation are prone to weaken the accurate transmission of cultural information. In addition, there are differences in audience knowledge background, media usage habits and platform distribution mechanism, and the acceptance effect of similar cultural content in different communication situations often fluctuates obviously, and the problems of insufficient content adaptation and unstable communication efficiency coexist. It can be seen that the key to the external communication of traditional Chinese culture is not only the adequacy of communication resources, but also how to accurately identify, structured organization and targeted optimization of cultural content, so as to achieve a higher level of understanding, recognition and acceptance in the cross-cultural context.

Around this topic, domestic and foreign research has formed a certain accumulation. The research on the international communication of traditional culture mainly focuses on the interpretation of cultural values, the construction of communication paths, the transformation of media forms, the remodeling of narrative mechanisms and the logic of audience acceptance, etc., which provides an important basis for revealing the content characteristics and operation rules of the external communication of traditional culture. The research on content optimization emphasizes the role of text organization, information configuration, expression strategy and scene adaptation in shaping the communication effect. At the same time, the rapid development of natural language processing, information retrieval and semantic computing provides a new technical path for cultural content research. Methods such as word embedding, sentence embedding, context embedding, semantic similarity calculation, multilingual representation learning and cross-lingual alignment can characterize the topic structure, context correlation and potential meaning relationship of the text from the distributed semantic level, and provide computational support for the recognition, comparison, aggregation and reorganization of complex cultural texts. However, there is still an obvious separation in existing research: cultural communication research focuses on meaning interpretation and communication phenomenon analysis, and pays insufficient attention to the computational expression of text semantic structure. Computer technology research focuses on model performance and task indicators, and has limited involvement in the cultural load, context dependence and cross-cultural interpretation complexity of traditional cultural content. Based on this, this paper intends to construct a content vectorization representation and optimization model for the external communication of Chinese traditional culture. With the help of text embedding, semantic correlation calculation, topic clustering and cross-language semantic mapping, the core concepts, narrative structure and value semantics in the content of traditional culture communication are vectorized. Furthermore, the differences in cultural identification, semantic integrity and audience adaptation are identified, and the corresponding content optimization paths are proposed, expecting to provide an analysis framework with both cultural interpretation power and technical operability for the research on international communication of traditional culture.

2 Theoretical basis and technical support

2.1 The content characteristics of Chinese traditional culture's external communication

The external communication content of traditional Chinese culture is not a general information text, but a compound semantic object with cultural load, symbol density and context dependence. Its core features are mainly reflected in the deep embedding of value expression, the hierarchical organization of narrative structure, the high-density bearing of cultural symbols and the complex adaptation of context transformation. As shown in Table 1, the content of traditional cultural communication shows strong structural characteristics in the four dimensions of value, narrative, symbol and context, which also determines that the content optimization cannot stay at the surface rhetorical adjustment, but should enter the level of semantic representation, relationship recognition and cross-cultural mapping. Cui et al. (2023), after studying the social media communication of cultural heritage institutions, proposed that the organization mode, interactive information configuration and contextualized expression of cultural content would significantly affect the audience's participation [1], indicating that the external communication of traditional culture first has a distinct value expression orientation, and its content is not simply stating facts. Instead, it conveys cultural identity, aesthetic ideas and social meaning through selective coding. Shim et al. (2024) proposed that digital narrative in cultural heritage communication is essentially the structural reconstruction of value information in the virtual environment [2], indicating that traditional cultural content usually relies on plot advancement, scene embedding and progressive meaning to form a narrative chain in the communication process, rather than presenting cultural elements in isolation.

In terms of cultural symbols, Lopez-Mugica et al. (2024) pointed out after studying large-scale international communication events that the visibility, combination mode and platform presentation logic of cultural symbols jointly affect the construction effect of national image [3]. Robinson-Jones (2024), after studying the linguistic landscape of multilingual museums, proposed that linguistic symbols, visual signs and spatial texts together constitute the explicit interface of cultural meaning [4]. This shows that the symbol system in the external communication of traditional culture has the characteristics of multi-modal coupling, which often includes the collaborative expression of language, image, space and ritual marks. After studying sensory narrative guide, Chan et al. (2025) proposed that the design of narrative nodes, the organization of sensory clues and the arrangement of spatial sequence can enhance the audience's immersion understanding [5], further indicating that the content of traditional cultural communication has significant sequential narrative characteristics. Yi et al. (2025), after studying the influence of digital experience on the communication of intangible cultural heritage, pointed out that the emotional activation mechanism can improve the audience's understanding, identification and sharing intention [6], indicating that the communication content also has the attribute of emotional trigger. After studying the transformation of intangible cultural heritage from the perspective of fashion communication, Xie et al. (2025) proposed that traditional cultural elements need to be restructured in modern context and re-encoded in communication scenes to improve cross-cultural acceptance [7]. Therefore, the optimization of traditional culture communication content is essentially a process of semantic reconstruction, symbol screening and context adaptation of texts with high cultural load.

Table 1: The basic characteristics of the external communication content of traditional Chinese culture

Feature Dimension	Main Manifestation	Technical Implication
Value Expression	Emphasizes cultural identity, ethical values, aesthetic spirit, and meaning transmission	Requires the identification of deep semantics and value themes
Narrative Structure	Relies on plot organization, scene progression, and progressive meaning construction	Requires modeling of textual sequence relationships and thematic chains
Cultural Symbols	Includes linguistic signs, visual imagery, spatial markers, and ritual elements	Requires multimodal symbol extraction and associative representation
Contextual Transformation	Involves cross-lingual transcription, cross-cultural interpretation, and scenario adaptation	Requires semantic mapping and communication adaptation optimization
Emotional Activation	Stimulates understanding, identification, and willingness to share through experience design	Requires characterization of emotional features and their association with audience responses

2.2 Foundations of Vector Algorithms and Semantic Representation

Vector algorithm and semantic representation technology provide the core method foundation for the computational modeling of complex text content. The basic idea is to map discrete language units into continuous vector space, and describe semantic relations by means of distance, direction and distribution relations. Li et al. (2022) proposed that the universal sentence representation method can integrate lexical, syntactic and contextual information through a unified vector space, thereby improving the consistency and transferability of text semantic expression [8]. This view shows that sentence vectors can not only compress text information, but also provide a stable representation basis for cross-text comparison and content matching. After studying the semantic retrieval model, Guo et al. (2022) pointed out that the retrieval method based on vector representation has shifted from keyword matching to semantic recall and deep representation learning [9], indicating that the core of text similarity judgment has shifted from word overlap to semantic proximity calculation. After studying the evolution path of word embedding, Incitti et al. (2023) proposed that semantic representation technology has been extended from early static word embedding to contextualized representation, multi-granularity embedding and cross-modal representation system [10], so that word embedding, sentence embedding and text embedding can support complex semantic modeling at different levels. Apidianaki (2023) proposed that the research on word sense representation has gradually shifted from word type modeling to word sense interpretation under context conditions [11], which means that the phenomena of polysemy, context dependence and cultural metaphor in traditional cultural texts can be more accurately depicted by contextualized embedding.

At the level of specific tasks, da Costa et al. (2023), after studying embedding methods in text classification, pointed out that vector representation can significantly improve the semantic discrimination ability of classification tasks through feature compression and semantic aggregation [12]. de Andrade et al. (2023) further proposed that high-quality context embedding can enhance the separability between categories, so that the representation space

itself has stronger structural discrimination ability [13]. This provides a technical basis for the content clustering analysis, topic identification and difference comparison of traditional culture communication in the following paper. Meanwhile, Wang et al. (2024), after studying the application of BERT in information retrieval, pointed out that pre-trained language models can improve the accuracy of semantic matching through deep context modeling, and provide a unified representation framework for similarity calculation, ranking optimization, and content recommendation [14]. Therefore, word embedding, sentence embedding, text embedding, semantic similarity calculation, clustering analysis and dimensionality reduction representation are not isolated method modules, but together form a semantic computing chain: Firstly, text encoding is completed by embedding representation, and then semantic proximity is depicted by similarity calculation, and then topic structure recognition and representation space visualization are realized by clustering and dimension reduction, which lays a technical foundation for vectorization modeling and optimization of traditional culture communication content.

2.3 Semantic Mapping mechanisms in Cross-language propagation

The translation of Chinese cultural texts is not a simple process of language conversion, but a complex mapping process involving concept reorganization, semantic projection and cultural interpretation. Since traditional Chinese cultural texts often contain historical context, ethical category, aesthetic imagery and metaphorical expression, problems such as semantic loss, cultural discount and discourse deviation are prone to occur in cross-language transfer. Qin et al. (2025) proposed that one of the core challenges of multilingual large language models is the consistency and transfer stability of semantic representation between different languages [15]. Xu et al. (2025), after studying the corpus, alignment and bias of multilingual models, pointed out that the lack of high-quality alignment of cross-lingual representations can easily lead to semantic center drift and cultural information weakening [16]. This conclusion shows that the concepts with high cultural load in Chinese cultural texts, such as "li", "Tao" and "Hehe", are often difficult to maintain the original meaning boundaries and value orientations if they are only replaced by words in foreign translations.

From the perspective of technical mechanism, semantic mapping in cross-language propagation is essentially to project source language texts and target language texts into a shared vector space, and maintain semantic proximity through representation alignment. Pallucchini et al. (2025) proposed that the key to cross-lingual contextualized representation alignment is to reduce the structural bias in different language encoding Spaces, so that semantic equivalence units can achieve neighbor aggregation in a unified vector space [17]. After studying the cross-lingual representation of low-resource languages, Fernando and Ranathunga (2025) pointed out that the cross-lingual stable representation ability of cultural names, institutional terms and entity concepts can be improved through entity mask and continuous alignment training [18]. After studying the task of cross-lingual and cross-temporal summarization, Zhang et al. (2024) proposed that cross-lingual content compression and reconstruction should maintain both topic consistency and semantic continuity, otherwise it will easily lead to narrative focus deviation [19]. Naorem et al. (2024) proposed that the method based on linear orthogonal mapping and graph structure constraints can improve the local neighborhood preservation ability in cross-lingual embedding space [20]. As shown in Figure 1, cross-lingual semantic mapping can be summarized as a technical chain of "Chinese cultural text encoding -- source language vector representation -- shared semantic space alignment -- target language semantic reconstruction -- dissemination content optimization". It can be seen that cross-lingual vector space alignment not only helps to

reduce semantic loss and cultural discount in translation communication, but also provides a computable basis for semantic matching, expression reorganization and international communication adaptation of traditional cultural content in the following text.

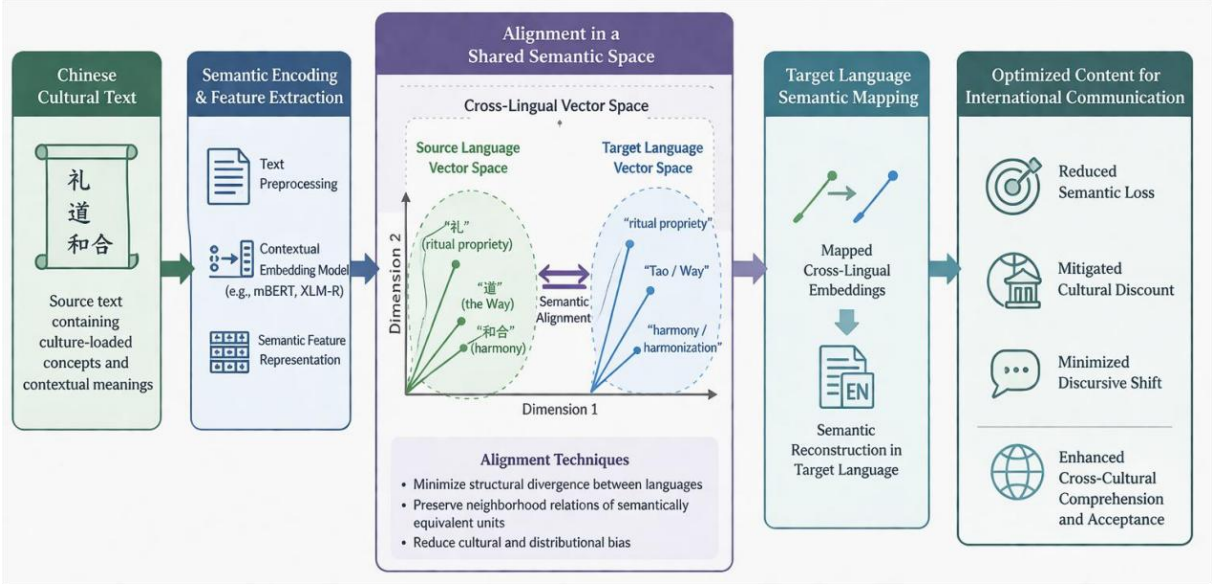

Figure 1: Schematic representation of the semantic mapping mechanism in cross-language propagation

3 Study design and model construction

3.1 Construction of corpus of external communication of traditional culture

In order to ensure the computability of subsequent vectorization modeling and content optimization analysis, this paper firstly constructs a multi-source heterogeneous corpus for the external communication of Chinese traditional culture. The corpus sources mainly include English translations of traditional classical books and multilingual translation texts, cultural communication texts on international communication platforms, external publicity materials of cultural institutions, introduction texts of overseas museums or cultural projects, and thematic web pages and digital guide materials for international audiences. In the collection stage, computer technologies such as web crawler, API interface crawling, HTML structure parsing and batch text extraction are used to gather data from different sources. Let the original corpus set be:

$$\mathcal{D}_0 = \bigcup_{k=1}^{K} \mathcal{S}_k \tag{1}$$

Here, Sk represents the KTH source text subset, and k is the number of data source categories. To control the deviation of sample distribution, the source weight vector is further defined as follows.

$$w = (w_1, w_2, \dots, w_K), \quad \sum_{k=1}^{K} w_k = 1 \tag{2}$$

In order to maintain a relative balance between classic texts, platform texts and institution texts in the subsequent training set.

The original text needs to complete coding unification, format cleaning and noise elimination before entering the library. Let a single text be xi, and its cleaning result is denoted as follows.

$$\tilde{x}_i = f_{norm}(f_{html}(f_{enc}(x_i))) \tag{3}$$

fenc stands for character encoding standardization, fhtml stands for tag and script stripping, and fnorm stands for Unicode normalization, punctuation normalization, and whitespace compression. In order to exclude excessively short text, directory text and abnormal fragments, a length filter function is defined:

$$g(\tilde{x}_i) = \begin{cases} 1, & L_{min} \le |\tilde{x}_i| \le L_{max} \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

Only the samples that satisfy the threshold constraint are kept. Considering that the corpus contains Chinese original text and multi-language translation, this paper introduces a language recognition model to determine the language of the text:

$$\hat{y}_i = \arg\max_{c \in \mathcal{C}} P(c \mid \tilde{x}_i) \tag{5}$$

where C is the language set and y^i is the predicted language label. For cross-lingual parallel texts, the alignment pairs are further established by title matching, paragraph anchor matching and sentence vector nearest neighbor search. Let the alignment score between Chinese sentence ui and foreign sentence vj be:

$$A_{ij} = \lambda_1 \, Sim_{sem}(u_i, v_j) + \lambda_2 \, Sim_{meta}(u_i, v_j) + \lambda_3 \, Sim_{pos}(u_i, v_j) \tag{6}$$

Here, Simsem is semantic similarity, Simmeta is metadata consistency, Simpos is location adjacency, and λ1+λ2+λ3=1.

For duplicate removal, a near-duplicate detection mechanism based on text embeding is used. Let the text vectors be ei and ej, and their cosine similarity is defined as follows:

$$\cos(e_i, e_j) = \frac{e_i^T e_j}{\|e_i\| \, \|e_j\|} \tag{7}$$

When cos(ei,ej)>τ, the two samples are judged as duplicate or nearly duplicate samples to avoid the interference of template publicity text on semantic distribution. In the corpus annotation stage, a multi-label annotation system is established around the theme category, value expression, cultural symbol, communication scene and target language. Let the label vector for the ith text be:

$$z_i = (z_{i1}, z_{i2}, \dots, z_{im}), \quad z_{ij} \in \{0,1\} \tag{8}$$

Where, m is the total number of tags. In order to evaluate the consistency of annotation, the consistency coefficient is introduced as follows:

$$\kappa = \frac{p_o - p_e}{1 - p_e} \tag{9}$$

Here, po is the observed agreement rate and pe is the random agreement rate. Only when κ reaches a preset threshold, the annotation result can enter the formal corpus.

In the preprocessing stage, we use word segmentation, sentence segmentation, stop word filtering, named entity recognition and term normalization to form input units suitable for vector representation learning. The final corpus is denoted as:

$$\mathcal{D} = \{(x_i, \hat{y}_i, z_i, m_i)\}_{i=1}^{N} \tag{10}$$

Among them, mi represents metadata features such as source, time, platform, and language. As shown in Figure 2, the corpus process constructed in this paper consists of five steps: "multi-source collection, parsing and normalization, cleaning and duplication removal, labeling and alignment, modeling and preprocessing", which not only ensures the cultural integrity of traditional cultural communication texts, but also provides high-quality data basis for subsequent text embedding, semantic mapping and content optimization model design.

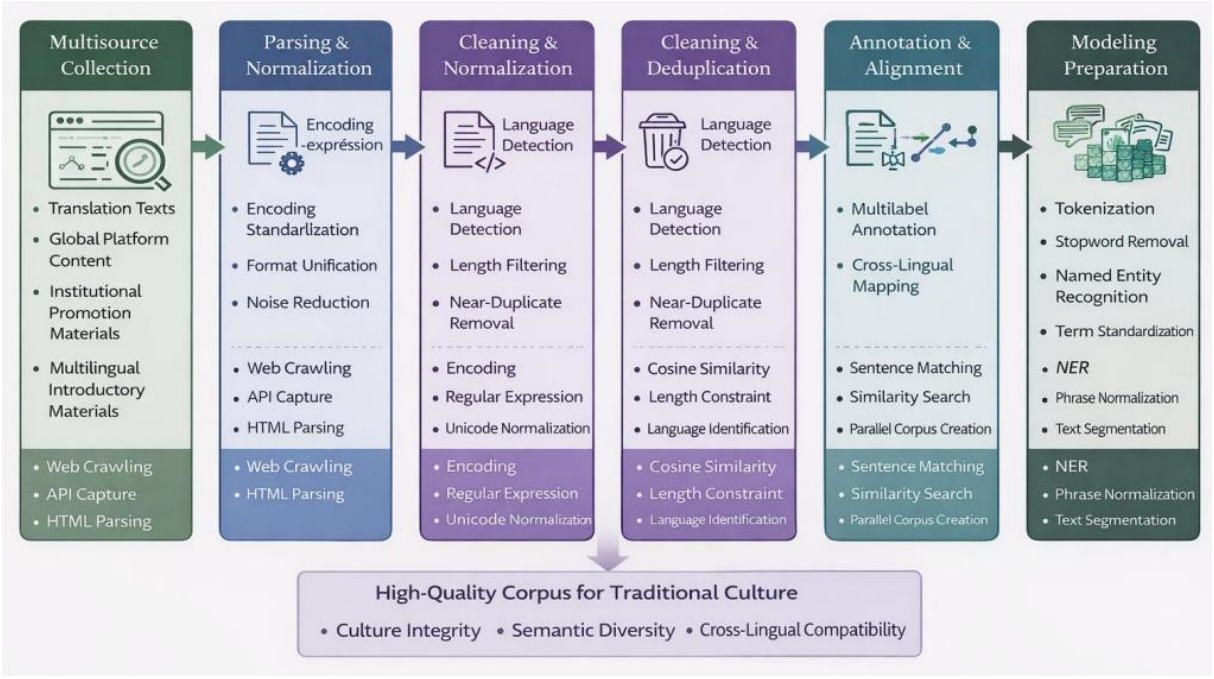

Figure 2: Illustration of the semantic mapping mechanism in cross-language propagation

3.2 Vectorized representation of the spread content

The vectorized representation of communication content is the key link to transform the external communication text of traditional culture from discrete language symbols to computable semantic structure. Considering the characteristics of traditional cultural communication texts, such as dense terminology, continuous narrative, obvious topic hierarchy and deep implicit cultural meaning, this paper does not adopt a single granularity representation method, but constructs a multi-level representation system composed of word level, sentence level and topic level, so as to realize the joint modeling of high-frequency topics, core semantics and cultural meaning. As shown in Figure 3, the system follows the basic path of "original text input-word-level encoding-sentence-level aggregation-topic-level abstract-multi-level fusion output", which not only retains fine-grained cultural word

information, but also takes into account sentence-level semantic organization and topic-level meaning generalization, so as to provide a unified representation basis for subsequent semantic matching, content optimization and cross-language mapping.

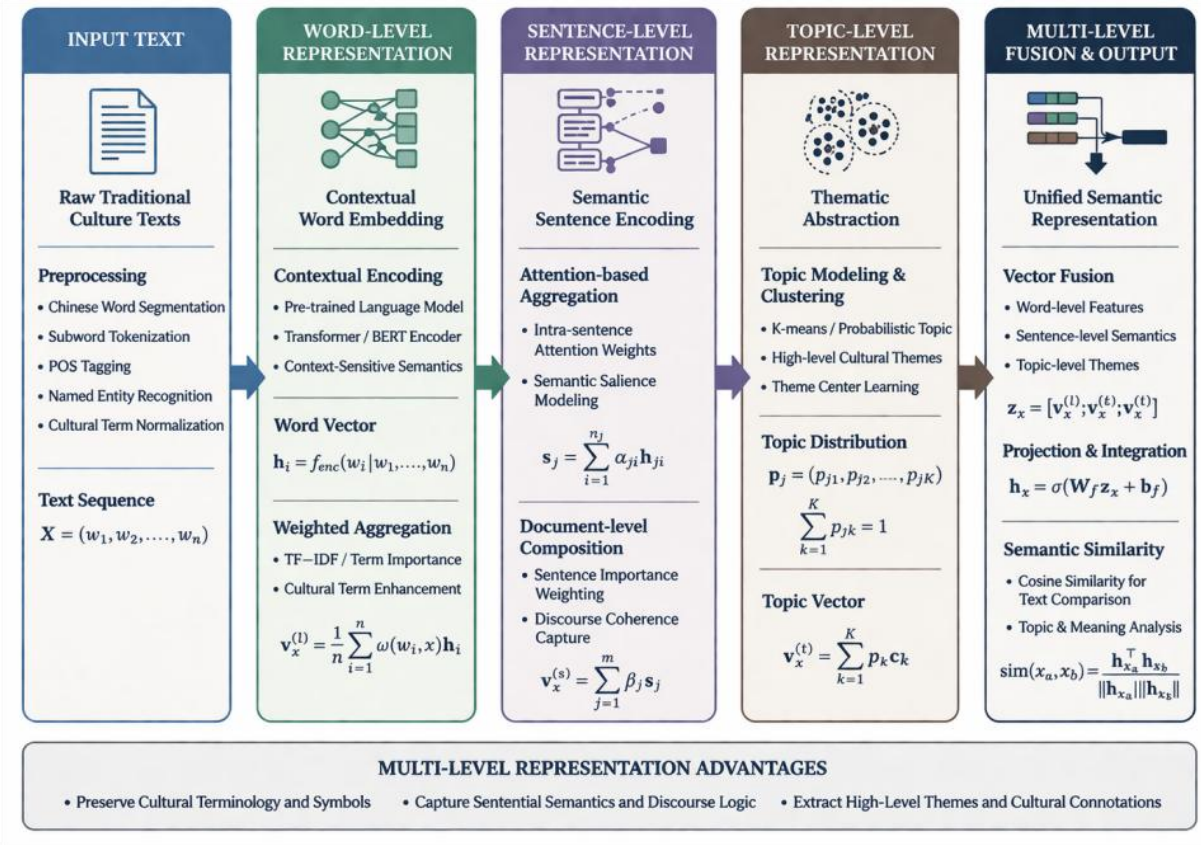

Figure 3:Multi-Level Vector Representation of Traditional Culture Communication Content

At the lexical level, the goal of vectorization is not to simply convert each word into a number sequence, but to retain the discriminative concept information in traditional cultural texts as much as possible. To this end, this paper firstly performs word segmentation, sub-word segmentation, part-of-speech tagging, named entity recognition and term normalization on the text, and distinguishes culture-loaded words such as "li", "le", "tao", "he", "Jianguo" and "unity of nature and man" from general function words. Each term is then mapped into a continuous vector space using pre-trained word embeddings or contextualized embedding models. If the text sequence is denoted as:

$$\mathrm{X} = (\mathrm{w}_1, \mathrm{w}_2, \ldots, \mathrm{w}_\mathrm{n}) \tag{11}$$

The vector representation of the ith term is as follows.

$$\mathrm{h}_\mathrm{i} = \mathrm{f}_{\mathrm{enc}}(\mathrm{w}_\mathrm{i} \mid \mathrm{w}_1, \ldots, \mathrm{w}_\mathrm{n}) \tag{12}$$

Here, fenc represents the representation function of the context-based encoder. Compared with the traditional static word embedding, this method can more effectively distinguish the meaning differences of cultural concepts in different contexts. For example, "li" has different semantic centers in different contexts, such as etiquette, institution, and ethical order. It is difficult to accurately distinguish "li" by only relying on word information, while the contextualized representation can preserve this difference through the antecedents and

posttexts. In order to further highlight the importance of cultural keywords, the word vector should be weighted by combining word frequency, term saliency and cultural symbol density in the vocabulary layer, so that high-value cultural words occupy a higher weight in the overall representation.

At the sentence level, the focus of the model shifts from "word meaning" to "semantic unit organization". Traditional cultural communication texts are usually not isolated statements, but organized by means of explanation, extension, comparison and narration. Many cultural meanings can only be fully recognized at the sentence level or even sentence group level. Therefore, this paper takes the sentence as the basic semantic aggregation unit, performs attention aggregation on the output of the lexical layer, and constructs a sentence vector that can represent the gist within the sentence and the cohesion relationship between sentences. If the JTH sentence consists of several word vectors, its representation can be summarized as follows.

$$s_j = \sum_{i=1}^{n_j} \alpha_{ji} h_{ji} \tag{13}$$

Here, αji represents the contribution weight of the ith word to the semantics of the sentence. This weight is not fixed, but dynamically adjusted according to contextual relevance, syntactic position and semantic saliency. In this way, terms that assume the functions of value expression, cultural interpretation and narrative turning point can obtain higher influence in the sentence vector. Furthermore, the sentence level should not only depict the meaning of a single sentence, but also identify the logical advancement relationship between sentences, such as the three-stage structure of "concept propotion-cultural interpretion-reality transformation", or the chain structure of "traditional meaning - international expression - communication purpose". Through the continuous aggregation of sentence vectors, the model can more accurately grasp the narrative organization mode and the progressive logic of meaning inside the communication text.

At the topic level, we further extract higher-level topic structures from sentence-level semantic units. In the external communication texts of traditional culture, many surface expressions may be different, but they may focus on the same cultural theme, such as etiquette norms, ethical order, natural aesthetics, family and country concepts, and intangible cultural heritage skills. If the model only stays at the word level or sentence level, it is easy to identify the local semantic proximity, but it is difficult to grasp the overall distribution of the full text in the topic space. Therefore, this paper further maps sentence vectors into topic vectors by topic clustering, topic mixture estimation and topic center modeling. Suppose there are K latent topics in the corpus, then the topic representation of the text can be simplified as follows.

$$v_x^{(t)} = \sum_{k=1}^{K} p_k c_k \tag{14}$$

Here, pk represents the probability that the text belongs to the KTH topic, and ck represents the center vector of the topic. The role of topic level representation is not only to discover high-frequency topics, but also to identify structural connections between different cultural concepts. For example, "li" and "he" may be far apart at the syntactic level, but they may belong to the same thematic cluster of ethical order and social coordination in the thematic space. With the help of the topic-level representation, the model can integrate the

scattered local semantics into a more explanatory cultural topic framework.

After the three levels of representation are completed, multi-level fusion is also required to obtain a unified propagation content vector. The reason is that the lexical layer is better at recognizing cultural terms and symbols, the sentence layer is better at expressing semantic continuity and narrative structure, and the topic layer is better at summarizing global meaning and high-level cultural orientation. Without the fusion mechanism, these three types of information will be scattered from each other, and it is difficult to form a stable and usable overall semantic representation. Therefore, this paper adopts a multi-layer vector fusion strategy to map the three types of representations into the same semantic space and generate a unified content vector. The semantic proximity between texts is measured by cosine similarity:

$$\mathrm{sim}(x_a, x_b) = \frac{h_{x_a}^{\top} h_{x_b}}{\|h_{x_a}\| \, \|h_{x_b}\|} \tag{15}$$

hxa and hxb are the fused text representations. This index can not only be used to identify similar cultural topic texts, but also be used to detect the semantic consistency between different translation versions. It can also be used as a basis for subsequent content optimization and expression reorganization.

In general, the multi-level vector representation method constructed in this paper is not a general text coding process, but a special representation framework designed for the characteristics of "prominent cultural terms, strong narrative interpretation, complex theme structure, and deep semantic level" of the external communication text of traditional culture. The core advantage of the proposed method is that it can extract sentence-level semantic relations and topic-level cultural meanings while preserving the details of cultural terms, so as to realize the unified modeling of high-frequency topic identification, core semantic extraction and cultural meaning discovery.

3.3 Content optimization Model based on vector algorithm

After completing the multi-level vector representation of communication texts, this paper further constructs a content optimization model for the external communication of traditional Chinese culture, so as to realize the linkage processing of content screening, semantic matching, topic aggregation and communication expression optimization. The model propagates a set of candidate texts D={x1,x2,... xN} as input, the candidate content is structured and optimized based on computer technologies such as text embedding, vector retrieval, semantic clustering, keyword recombination and cross-language matching. The core idea is not to directly rewrite the original text on the surface, but to identify "which content is more worth preserving, which expressions are more suitable for reorganization, which topics need to be highlighted, and which translation texts are closer to the target communication context" in the semantic space, and then complete the content optimization according to the vector relationship.

In the content screening link, the model focused on solving the problems of information redundancy, repeated expression and scattered value focus in the external communication text of traditional culture. Let the fusion vector of text xi be denoted as hi and the target propagation topic vector as q, then the relevance between text and target propagation intention is defined as follows.

$$r_i = \frac{h_i^{\top} q}{\|h_i\| \|q\|} \tag{16}$$

Among them, the larger ri is, the closer the text content is to the preset propagation topic. In order to avoid content simplification only based on similarity, this paper further introduces information density coefficient ρi and cultural feature strength coefficient ηi to construct a comprehensive screening score:

$$F_i = \alpha r_i + \beta \rho_i + \gamma \eta_i, \quad \alpha + \beta + \gamma = 1 \tag{17}$$

Here, ρi is used to measure the effective semantic carrying capacity in a unit text length, and ηi is used to characterize the concentration of cultural terms, value expressions, and symbolic concepts. Through the scoring mechanism, the model can ensure the consistency of the topic while preferentially retaining the communication units with complete cultural information and high expression density.

In the semantic matching step, the model mainly deals with the synonymy identification and semantic alignment between the source text, the candidate text and the target propagation expression. For any two texts xi and xj, the semantic matching score is defined as follows.

$$S_{ij} = \cos(h_i, h_j) = \frac{h_i^\top h_j}{\|h_i\| \|h_j\|} \tag{18}$$

This index can be used to determine whether there is expression duplication, narrative approximation, or cultural meaning overlap between different communication materials. For further cross-language retrieval, if the Chinese text is represented as $h_i^{(zh)}$ and the foreign text is represented as $h_j^{(en)}$, the cross-language matching score is written as follows.

$$S_{ij}^{(cl)} = \frac{\left(h_i^{(zh)}\right)^\top W h_j^{(en)}}{\|h_i^{(zh)}\| \, \|W h_j^{(en)}\|} \tag{19}$$

Where W is the cross-lingual mapping matrix, which is used to project different language texts into a shared semantic space. Based on this mechanism, the model can retrieve the foreign language expressions closest to the Chinese cultural content in the multilingual corpus, and provide candidate text support for subsequent communication adaptation and translation optimization.

In the topic aggregation step, the model identifies the high-frequency topics and their internal structural relationships in the propagation content by vector clustering. Let the set of all text vectors be {h1,h2,... hN}, K-means clustering is used to construct topic clusters, and the objective function is as follows.

$$J = \sum_{i=1}^{N} \sum_{k=1}^{K} z_{ik} \|h_i - c_k\|^2 \tag{20}$$

Here, z k∈{0,1} indicates whether the text xi belongs to the KTH topic cluster, and ck is the KTH topic center. Through topic aggregation, the model can identify core communication topics such as "etiquette norms", "ethical order", "natural aesthetics", "intangible cultural heritage skills" and "family and country concepts", and further calculate the topic weights:

$$\pi_k = \frac{1}{N}\sum_{i=1}^{N} z_{ik} \tag{21}$$

In order to reflect the coverage degree of a certain topic in the communication corpus. This process not only helps to find high-frequency topics, but also helps to identify over-repeated topics and relatively missing topics in communication, providing a basis for the structural balance of communication content.

In the keyword reorganization and expression optimization step, the model does not directly generate the full text, but reconstructs and ranks the high-value keywords and semantic fragments. Let the word vector of the MTH keyword in the text xi be kim, and its comprehensive importance is defined as follows.

$$\omega_{im} = \lambda_1 \, tfidf_{im} + \lambda_2 \, att_{im} + \lambda_3 \, top_{im} \tag{22}$$

Among them, tfidfim represents the statistical significance of a keyword, attim represents its attention weight in the context encoding, and topim represents its contribution to the current topic cluster. According to the ranking results of ωim, the model can extract a set of keywords with both cultural representation and communication effectiveness, and reorganize them according to the theme consistency and expression order. If the expression sequence after recombination is denoted as Y=(y1,y2,... ,yT), then its optimization objective can be expressed as follows.

$$\max \ \Phi(Y) = \mu_1 Rel(Y, q) + \mu_2 Coh(Y) + \mu_3 Cult(Y) \tag{23}$$

Among them, Rel(Y,q) measures the relevance between the output expression and the target communication topic, Coh(Y) measures the internal semantic coherence of the text, and Cult(Y) measures the degree of cultural connotation retention. This objective function shows that communication expression optimization is not a single pursuit of language conciseness or text fluency, but a balance between topic-relevance, logical coherence and cultural fidelity.

In order to improve the overall discrimination ability of the model, this paper further uses the idea of learning to rank to train the optimization results. Assuming that positive samples are high-quality propagated texts xi+ and negative samples are low-quality or redundant propagated texts xi−, the margin loss function can be constructed as follows.

$$\mathcal{L} = \sum_{i=1}^{N} \max(0, \delta - F(x_i^+) + F(x_i^-)) \tag{24}$$

Here, δ is the minimum interval threshold. Through this loss function, the model can gradually learn which text is more in line with the requirements of topic concentration, cultural integrity and expression adaptation in external communication, thus enhancing the stability of content optimization.

In general, the content optimization model based on vector algorithm constructed in this paper forms a technical chain of "vector representation - related screening - semantic matching - topic aggregation - keyword reorganization - expression optimization". Its advantage is that it can transform the abstract problems in traditional cultural communication into computable semantic relationship processing problems, and realize the transformation of communication content from experience revision to data-driven optimization with the help of computer technologies such as text embedding, similarity calculation, clustering analysis,

cross-language retrieval and learning to rank. This provides an operational model basis for the subsequent establishment of content evaluation system and optimization strategy for international communication scenarios.

3.4 Evaluation index and optimization mechanism

In order to ensure that the communication content optimization does not stop at empirical judgment, but is based on a quantifiable, comparable and iterative technical framework, this paper further constructs an evaluation index system for the external communication of Chinese traditional culture, and designs a content optimization mechanism on this basis. The system is supported by computer technologies such as semantic computing, cross-language representation learning, text quality assessment and user behavior analysis, and comprehensively evaluates the optimization results from five dimensions: semantic accuracy, cultural fidelity, cross-cultural understandability, communication adaptability and user feedback effect.

Table 2: Content optimization evaluation index system and its calculation basis

Evaluation Dimension	Core Meaning	Main Calculation Metrics	Related Computer Technologies	Optimization Effect
Semantic Accuracy	Optimize the degree of semantic consistency between the text and the original text	Vector similarity, keyword consistency rate	Text embedding, cosine similarity, semantic matching	Prevent core semantic deviation
Cultural Fidelity	The degree of preservation of cultural concepts, terms, and value implications	Cultural term coverage, concept alignment score	Term extraction, entity recognition, concept embedding	Prevent loss of cultural connotations
Cross-cultural Comprehensibility	The difficulty for international audiences to understand and interpret the text	Readability, clarity, ambiguity penalty	Language models, complexity analysis, target corpus matching	Improve interpretation efficiency
Dissemination Adaptability	The degree of match between the text and the platform, scenario, and audience	Scenario vector similarity, style matching degree	Platform profiling, style classification, vector retrieval	Enhance scenario applicability
User Feedback Effect	The actual dissemination response after content is published	Click-through rate, interaction rate, save rate, dwell time	Log analysis, behavior modeling, online evaluation	Verify optimization effectiveness

As shown in Table 2, different dimensions correspond to content semantic consistency, cultural information retention degree, target audience understanding difficulty, platform scene matching level and communication feedback performance, so as to form a multi-dimensional evaluation framework covering "text quality - communication adaptation - user response".

In the semantic accuracy dimension, the evaluation focuses on whether the optimized text maintains the core semantics of the original communication content. Let the original text vector be hx and the optimized text vector be hy, then the semantic accuracy can be defined as follows.

$$A_{sem} = \frac{h_x^T h_y}{\|h_x\| \, \|h_y\|} \tag{25}$$

This index is realized based on text embedding and cosine similarity calculation, which can effectively measure how well the optimized text preserves the original text in the overall meaning space. If keyword-level constraints are introduced, the local consistency check can be further set for high-weight cultural terms to avoid the situation that the overall similarity but the key concepts deviate.

Cultural fidelity emphasizes the retention level of traditional cultural load information in dissemination optimization. Suppose that there are M core cultural concepts in the text, the importance weight of cultural concept cm is ωm, and its retention degree in the optimized text is denoted as rm∈[0,1], then the cultural fidelity can be expressed as follows.

$$A_{cul} = \frac{\sum_{m=1}^{M} \omega_m \, r_m}{\sum_{m=1}^{M} \omega_m} \tag{26}$$

Among them, rm can be jointly determined by term matching, concept embedding similarity, or cross-lingual alignment score. The technical basis of the index lies in named entity recognition, cultural term extraction and concept vector matching, which can avoid the problems of weakening of cultural proper names, loss of value meanings or excessive simplification of symbolic system in the optimization process.

Cross-cultural intelligibility mainly measures the difficulty of interpretation and acceptance of the target text for the international audience. Because traditional cultural texts often contain historical allusions, ethical categories and abstract images, semantic consistency alone is not enough to ensure audience understanding. Therefore, this paper defines intelligability as the comprehensive result of readability, semantic clarity and ambiguity punishment. Let Read(y) denote the readability score for the target language audience, Clar(y) denote the semantic clarity of the text, and Amb(y) denote the ambiguity strength, then:

$$A_{com} = \lambda_1 \mathrm{Read}(y) + \lambda_2 \mathrm{Clar}(y) - \lambda_3 \mathrm{Amb}(y) \tag{27}$$

Here, λ1+λ2+λ3=1. The relevant calculation can be completed by the perplexity of the language model, the syntactic complexity, the density of ambiguous words and the matching degree of the commonly used corpus of the target language, so that the evaluation not only stays on the text form, but also implements the cross-cultural interpretation efficiency.

Propagation fitness focuses on the degree of match between the optimization results and the specific propagation scenario. When traditional cultural content is presented to different channels such as social media, institutional official websites, digital exhibitions, short video copywriting, there are obvious differences in length, style, rhythm and information density. Let the feature vector of the platform or target scene be p and the scene representation vector

of the optimized text be hy(p), then the propagation adaptability can be written as follows.

$$A_{adp} = \mu_1 \frac{(h_y^{(p)})^\top p}{\|h_y^{(p)}\| \, \|p\|} + \mu_2 \text{Form}(y) + \mu_3 \text{Style}(y) \tag{28}$$

Among them, Form(y) measures the degree of conformity with the formal characteristics such as length, structure, and keyword density, and Style(y) measures the degree of matching between the text style and the platform style. This dimension mainly relies on the platform text portrait, style classification model and vector retrieval technology to complete the evaluation.

The user feedback effect is used to verify whether the content optimization actually improves the dissemination performance. Let click rate, interaction rate, collection rate and stay TIME be CTR, ER, SAVE and Time respectively, then the feedback score can be defined as follows.

$$A_{fb} = \sigma(\rho_1 \text{CTR} + \rho_2 \text{ER} + \rho_3 \text{SAVE} + \rho_4 \text{TIME}) \tag{29}$$

where $\sigma(\cdot)$ is the normalization function and ρi is the index weight. This part relies on user behavior log analysis, click stream modeling and online feedback monitoring, which is the part closest to the real propagation effect in the evaluation system.

At the level of comprehensive evaluation, this paper unifies five types of indicators as an overall optimization objective function:

$$Q(y) = w_1 A_{sem} + w_2 A_{cul} + w_3 A_{com} + w_4 A_{adp} + w_5 A_{fb}, \quad \sum_{i=1}^{5} w_i = 1 \tag{30}$$

where Q(y) represents the overall quality score of the optimized text y. Based on this function, the content optimization mechanism can continuously improve the model parameters by learning to rank or updating iteratively. If the model parameters are denoted as θ, we can update them as follows:

$$\theta_{t+1} = \theta_t - \alpha \nabla_\theta \mathcal{L}(\theta_t), \quad \mathcal{L}(\theta) = -Q(y) + \lambda \|\theta\|^2 \tag{31}$$

This indicates that the model will continuously revise the content filtering, keyword recombination, semantic matching and expression optimization strategies according to the multi-dimensional evaluation results, thus forming a closed-loop mechanism of "evaluation-feedback-re-optimization".

In summary, the evaluation index system established in this paper does not understand text optimization as a single language polishing process, but regards it as a comprehensive optimization task with the joint effects of semantic preservation, cultural fidelity, understanding promotion, scene adaptation and feedback improvement.

4 Experimental results and analysis

4.1 Analysis of vector distribution characteristics of traditional cultural communication content

In order to test the ability of the constructed multi-level vector representation method to

describe the structure of traditional cultural communication content, this paper firstly maps the communication text in the corpus to a unified semantic space, and uses the dimensionality reduction projection method to visualize the high-dimensional vector, so as to observe the clustering distribution characteristics of different types of traditional cultural content in the vector space. The experimental results show that different content types present a relatively obvious cluster structure in the low-dimensional semantic space, which indicates that the joint representation system of word level, sentence level and topic level constructed in this paper can effectively distinguish the differences in theme semantics, narrative mode and cultural meaning of traditional cultural communication texts.

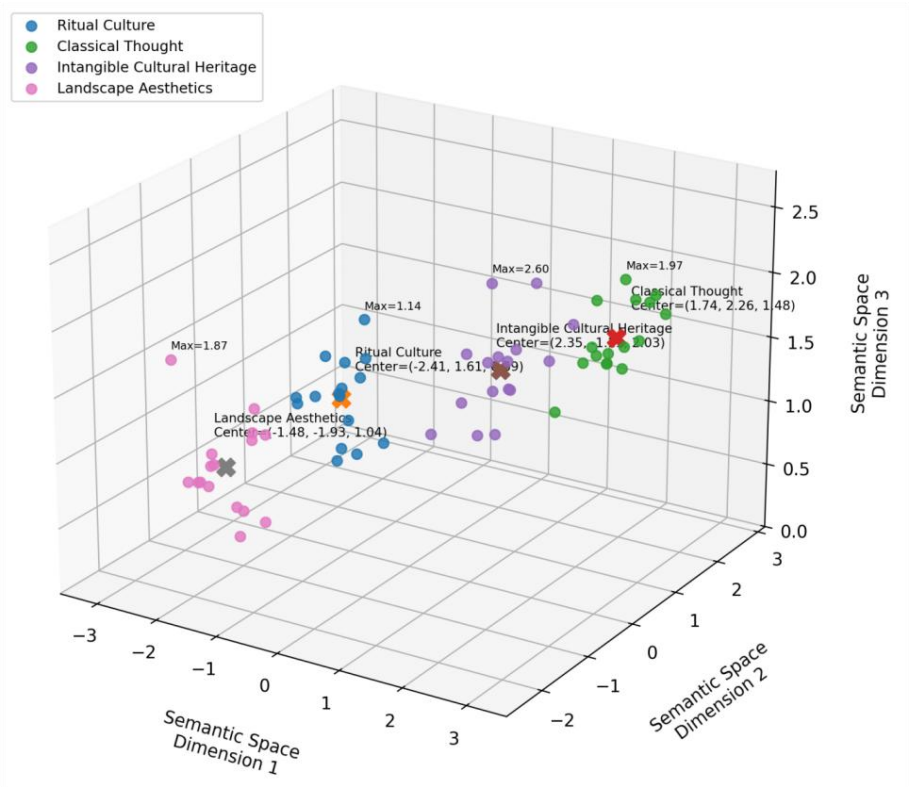

Figure 4: Three-Dimensional Scatter Distribution of Traditional Culture Communication Content in Semantic Vector Space

As shown in Figure 4, four types of texts, including etiquette culture, classical thought, intangible cultural heritage skills and landscape aesthetics, form relatively independent distribution areas in the three-dimensional semantic space. Among them, texts of intangible cultural heritage skills show higher agglomeration in the third dimension, indicating that such texts have stronger consistency in "skill process, cultural inheritance and experience expression". The classical ideological texts show a higher extension in the first and second dimensions, reflecting their stronger semantic tension in value interpretation, concept abstraction and explanation.

Further combining the mapping results of theme center and audience, it can be found that there is a significant structural correlation between cultural theme, narrative expression and communication audience. Figure 5 illustrates the distribution of alignment scores between different topics and different audience types.

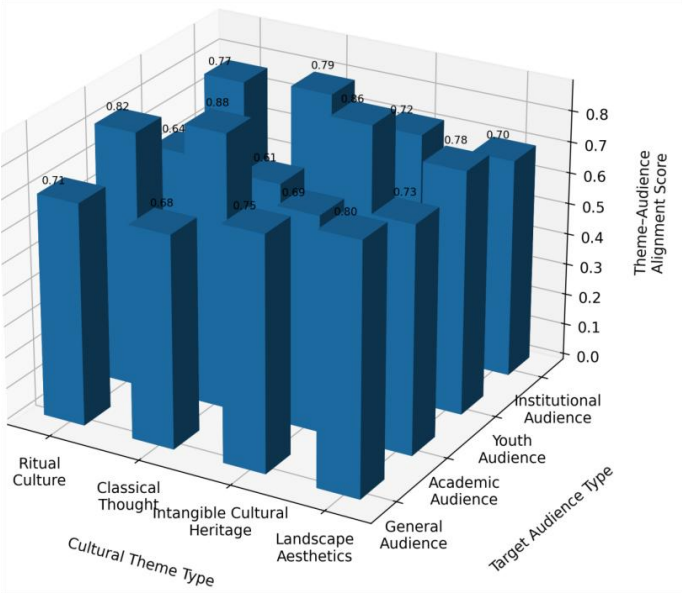

Figure 5: Three-Dimensional Bar Chart of Alignment Scores between Cultural Themes and Target Audiences

The results show that the classical thought theme has the highest matching degree in the academic audience, the intangible cultural heritage art theme has a relatively higher response score in the young audience, and the adaptation of the landscape aesthetic theme is more prominent in the mass audience. This indicates that there is no uniform optimal expression path for different traditional cultural themes in external communication, but it is influenced by theme attributes, narrative methods and knowledge background of target audiences. In other words, the topic distance in vector space not only reflects the semantic proximity between texts, but also reveals the correspondence between communication objects and expression strategies to a certain extent.

From the perspective of vector distribution quality, different content types have certain differences in clustering compactness and inter-class separation.

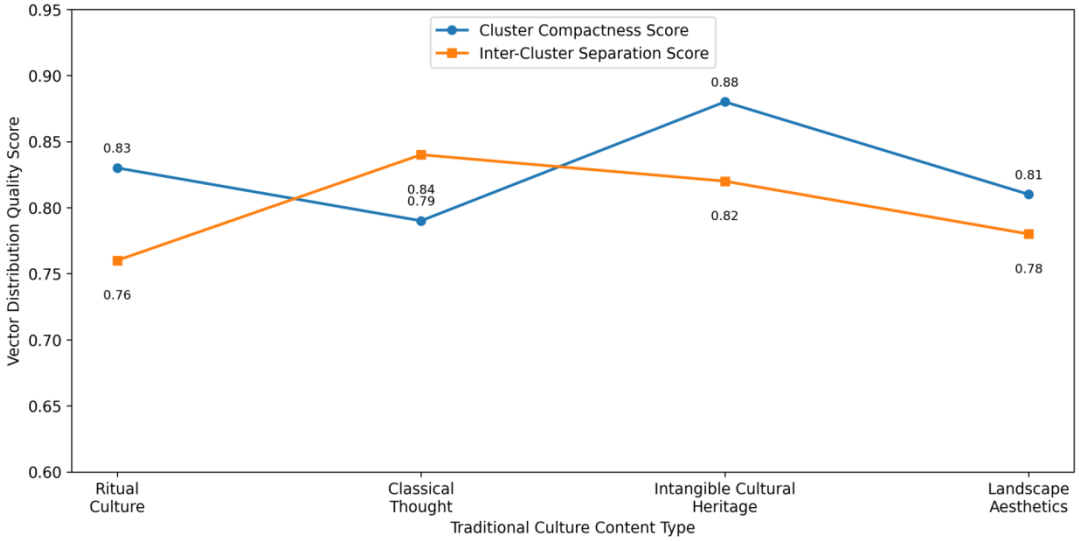

Figure 6: Vector Distribution Quality Scores across Different Types of Traditional Culture Content

As shown in Figure 6, the clustering compactness of intangible cultural heritage technical texts is the highest, reaching 0.88, indicating that it has strong pattern stability in terms of terminology system, scene description and technological process. The inter-class separation of classical ideological texts is relatively higher, reaching 0.84, indicating that although the internal expression forms of this type of text are relatively rich, it still has strong discrimination ability in distinguishing from other cultural topics. Etiquette culture and landscape aesthetic texts show a medium level of compactness and separation, indicating that their communication and expression are influenced by value narrative and image rhetoric to a certain extent, so the boundary in vector space is relatively flexible. In general, the proposed model can better reveal the clustering law of traditional cultural communication content in semantic space, and provide visual basis and structural support for subsequent content screening, topic aggregation and expression optimization.

4.2 Effect analysis of content optimization based on vector algorithm

In order to verify the actual role of vector algorithm in the improvement of traditional cultural communication content, this paper compares and analyzes the three dimensions of semantic consistency, communication clarity and cross-cultural fitness of the text before and after optimization. The results show that the overall performance of the optimized text is significantly improved. The semantic consistency was improved from 0.74 to 0.87, with an increase of 17.6%. The communication clarity was increased from 0.69 to 0.83, with an increase of 20.3%. The cross-cultural fitness increased from 0.66 to 0.81, with an increase of 22.7%.

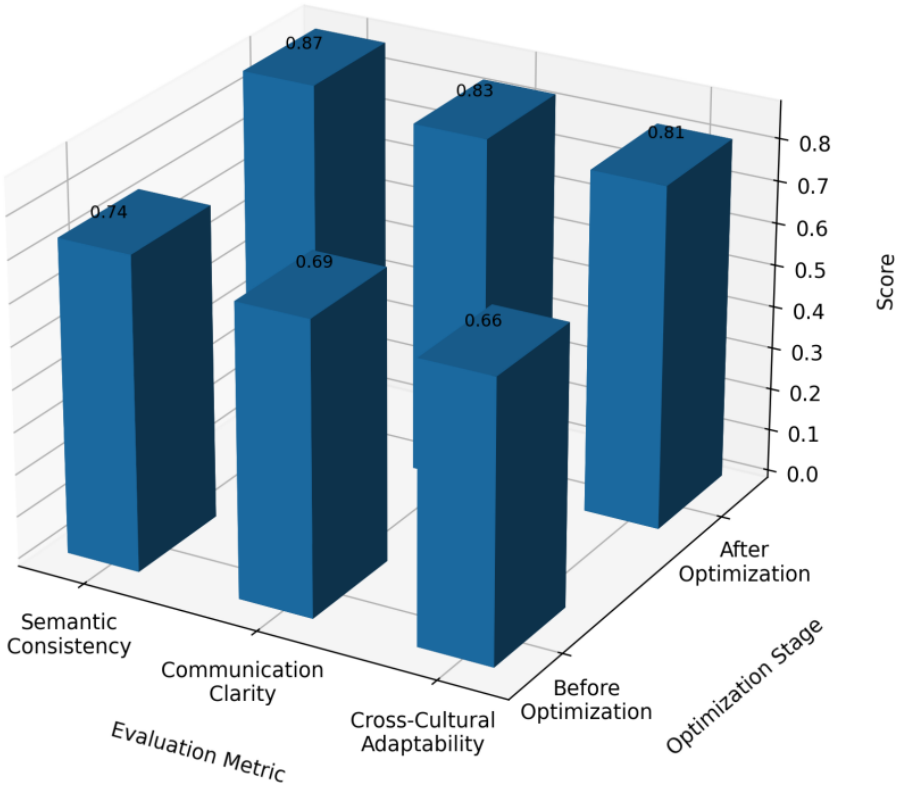

Figure 7: Three-Dimensional Bar Chart of Content Optimization Performance

As shown in Figure 7, each index after optimization is significantly higher than that before optimization, indicating that content screening, semantic matching and keyword

reorganization based on vector representation can effectively reduce expression redundancy and semantic deviation, and enhance the ability of text topic focusing and the stability of meaning transmission. Further, from the perspective of different content types, the comprehensive optimization scores of four types of texts, including etiquette culture, classical ideas, intangible cultural heritage skills and landscape aesthetics, increased from 0.70, 0.72, 0.68 and 0.71 to 0.84, 0.86, 0.82 and 0.85, respectively.

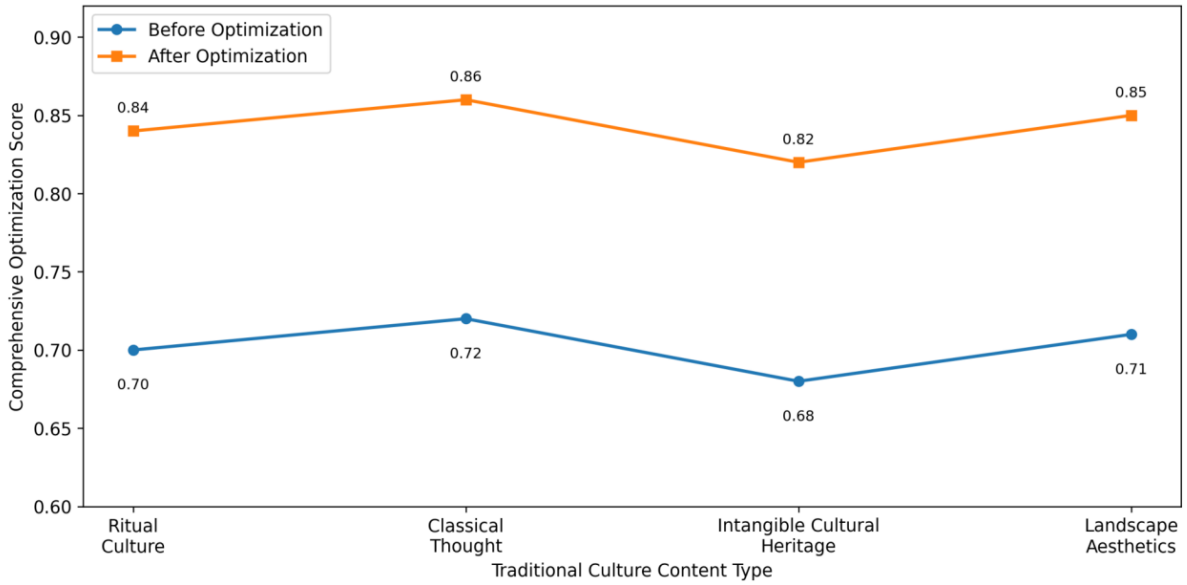

Figure 8: Changes in Comprehensive Scores before and after Optimization

As shown in Figure 8, all kinds of texts show a stable upward trend after optimization, and the improvement range of classical ideological texts is the most obvious, indicating that the vector algorithm can more effectively improve the quality of semantic organization and cross-cultural expression when dealing with cultural contents with dense abstract concepts and complex explanation paths. In general, the vector algorithm not only improves the semantic consistency of the text, but also improves the expression clarity and audience adaptation in international communication, which verifies its technical effectiveness in the content optimization of traditional cultural communication.

4.3 Comparison experiments of different models

In order to further verify the effectiveness of the proposed method in the content optimization task of traditional cultural communication, this paper conducts comparative experiments with the traditional keyword matching method, rule method and general text classification method. The comparison metrics mainly include semantic recognition accuracy, content recommendation accuracy, cross-cultural matching score and comprehensive performance score. Among them, the international communication content recommendation task is constructed based on the semantic similarity ranking results, and the semantic recognition task takes the manually labeled topic tags as the reference standard. Experimental results show that the model based on vector algorithm is superior to the comparison methods in all indicators. As shown in Table 3, the accuracy of traditional keyword matching method in semantic recognition is only 0.73, and the accuracy of content recommendation is 0.69, indicating that this method mainly relies on word surface overlap and is difficult to identify implicit semantic relationships. The rule method was improved to 0.78 and 0.74, respectively, but limited by the coverage of artificial rules, the stability was insufficient in the face of

complex cultural expression and cross-context conversion. The semantic recognition accuracy and recommendation accuracy of general text classification methods reach 0.83 and 0.79, respectively, which have certain classification ability, but the deep association mining between cultural concepts is still insufficient. In contrast, the semantic recognition accuracy, content recommendation accuracy and cross-cultural matching score of the proposed method reach 0.89, 0.86 and 0.84, respectively, and the comprehensive performance score reaches 0.87, which is 7.2%, 8.9%, 10.5% and 8.8% higher than that of general text classification methods. This shows that the vector algorithm can effectively capture the concept connection, narrative structure and topic aggregation relationship in traditional cultural texts through continuous semantic space modeling, and has stronger advantages in dealing with near-meaning expression, metaphorical meaning and cross-language mapping. In general, the method in this paper shows better performance in terms of semantic recognition accuracy, content recommendation ability and cross-cultural adaptation effect, which verifies the technical advantages of vector algorithm in the optimization of traditional culture external communication content.

Table 3: Comparative experimental results of different models

Model	Semantic Recognition Accuracy	Content Recommendation Accuracy	Cross-Cultural Matching Score	Overall Performance Score
Keyword Matching Method	0.73	0.69	0.65	0.69
Rule-Based Method	0.78	0.74	0.70	0.74
General Text Classification Method	0.83	0.79	0.76	0.80
Proposed Method	0.89	0.86	0.84	0.87

5 Conclusion and Prospect

Focusing on the problems of content homogeneity, semantic deviation and insufficient cross-cultural adaptation in the external communication of traditional Chinese culture, this paper constructs a research framework for communication content recognition, vector modeling and optimal output. The results show that the multi-level vector representation method based on vocabulary level, sentence level and topic level can better describe the core concepts, narrative structure and cultural meaning in traditional cultural communication texts. Different types of texts show a relatively clear clustering distribution in the semantic space, and the clustering compactness of intangible cultural heritage texts reaches 0.88. The inter-class separation of classical thought texts reaches 0.84, indicating that the constructed method has strong content recognition ability. In the content optimization experiment, the semantic consistency is improved from 0.74 to 0.87, the communication clarity is improved from 0.69 to 0.83, and the cross-cultural fitness is improved from 0.66 to 0.81, which indicates that the vector algorithm can effectively improve the topic focus, expression stability and international communication adaptation ability of the communication text. In the model comparison experiment, the semantic recognition accuracy, content recommendation accuracy and comprehensive performance score of the proposed method reach 0.89, 0.86 and 0.87, respectively, which are significantly better than keyword matching, rule method and general text classification method, indicating that it has strong technical advantages in traditional cultural content optimization. The theoretical value of this paper is that the international

communication research of traditional culture, digital humanities research and semantic computing methods are included in the same analysis chain, which expands the methodological boundaries of traditional culture communication research. The practical significance is that it provides a computable, comparable and optimized technical path for content screening, expression reorganization, topic aggregation and cross-language adaptation in the international communication of traditional culture. However, there are still some limitations in this paper, mainly reflected in the size of the corpus still needs to be expanded, the cross-language generalization ability needs to be improved, and the modeling of the deep structure of cultural semantics, metaphor mechanism and complex knowledge association is still not sufficient. Subsequent research can further introduce multi-modal communication data such as images, audio, and video, combine intelligent generation technology to improve the ability of content adaptive reconstruction, and enhance the ability of cultural concept correlation modeling and cross-language semantic reasoning through knowledge graph fusion, so as to promote the development of traditional culture external communication research to a higher level of intelligence and refinement.

References

[1] Cui T, Kumar P, Orr S A. Connecting characteristics of social media activities of a heritage organisation to audience engagement[J]. Digital Applications in Archaeology and Cultural Heritage, 2023, 28: e00253.

[2] Shim H, Oh K T, O'Malley C, et al. Heritage values, digital storytelling, and heritage communication: the exploration of cultural heritage sites in virtual environments[J]. Digital Creativity, 2024, 35(2): 171-197.

[3] Lopez-Mugica J, Whyke T W, White A, et al. The fluidity of soft power in YouTube's broadcast of the Beijing 2022 Winter Olympic Games[J]. Global Media and Communication, 2024, 20(3): 311-328.

[4] Robinson-Jones C. Linguistic landscapes of intangible cultural heritage museums representing minority languages: the case of the 'Gerhard Rohlfs' Museum of the Calabrian Greek Language[J]. Journal of Multilingual and Multicultural Development, 2024: 1-19.

[5] Chan C S, Shek K F, Lu Y, et al. Narratives come alive: Enhancing cultural heritage experiences through sensorial narrative-based guiding tours[J]. Journal of Heritage Tourism, 2025, 20(4): 487-507.

[6] Yi C, Huang J, Song L. Enhancing intangible cultural heritage dissemination through digital experience: an Affective Events Theory approach[J]. npj Heritage Science, 2025, 13(1): 438.

[7] Xie C, Lai F, Zhang J, et al. Transforming intangible cultural heritage in destinations: A fashion communication perspective[J]. Tourism Management, 2025, 110: 105161.

[8] Li R, Zhao X, Moens M F. A brief overview of universal sentence representation methods: A linguistic view[J]. ACM Computing Surveys (CSUR), 2022, 55(3): 1-42.

[9] Guo J, Cai Y, Fan Y, et al. Semantic models for the first-stage retrieval: A

comprehensive review[J]. ACM Transactions on Information Systems (TOIS), 2022, 40(4): 1-42.

[10] Incitti F, Urli F, Snidaro L. Beyond word embeddings: A survey[J]. Information Fusion, 2023, 89: 418-436.

[11] Apidianaki M. From word types to tokens and back: A survey of approaches to word meaning representation and interpretation[J]. Computational Linguistics, 2023, 49(2): 465-523.

[12] Da Costa L S, Oliveira I L, Fileto R. Text classification using embeddings: a survey[J]. Knowledge and Information Systems, 2023, 65(7): 2761-2803.

[13] de Andrade C M V, Belem F M, Cunha W, et al. On the class separability of contextual embeddings representations–or "the classifier does not matter when the (text) representation is so good!"[J]. Information Processing & Management, 2023, 60(4): 103336.

[14] Wang J, Huang J X, Tu X, et al. Utilizing bert for information retrieval: Survey, applications, resources, and challenges[J]. ACM Computing Surveys, 2024, 56(7): 1-33.

[15] Qin L, Chen Q, Zhou Y, et al. A survey of multilingual large language models[J]. Patterns, 2025, 6(1).

[16] Xu Y, Hu L, Zhao J, et al. A survey on multilingual large language models: Corpora, alignment, and bias[J]. Frontiers of Computer Science, 2025, 19(11): 1911362.

[17] Pallucchini F, Malandri L, Mercorio F, et al. Lost in alignment: A survey on cross-lingual alignment methods for contextualized representation[J]. ACM Computing Surveys, 2025, 58(5): 1-34.

[18] Fernando A, Ranathunga S. Linguistic entity masking to improve cross-lingual representation of multilingual language models for low-resource languages: A. Fernando, S. Ranathunga[J]. Knowledge and Information Systems, 2025, 67(11): 9905-9946.

[19] Zhang R, Ouni J, Eger S. Cross-lingual cross-temporal summarization: Dataset, models, evaluation[J]. Computational Linguistics, 2024, 50(3): 1001-1047.

[20] Naorem D, Singh S R, Sarmah P. Improving linear orthogonal mapping based cross-lingual representation using ridge regression and graph centrality[J]. Computer Speech & Language, 2024, 87: 101640.