Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 15, No. 3, 2026

G-T-ERNIE: Multi-Label Classifier with Text–Label Joint Modeling for Tourism Texts

Tuo Zhou, Busheng Li *

Jingdezhen Ceramic University, Jingdezhen, Jiangxi, China
* **Corresponding author**: Busheng Li (Email: 004267@jcu.edu.cn)

Abstract: With the rapid growth of user reviews on tourism platforms and the rising demand for intelligent services, multi-label classification has become essential for review retrieval and service optimization. Traditional methods struggle with the complex semantics and multi-label nature of reviews, leading to weak feature representation and lower accuracy. To address this issue, this study proposes a G-T-ERNIE-based multi-label text classification model. ERNIE is used to capture contextual semantics, TextCNN extracts local n-gram features, and a label co-occurrence graph with GCN models dependencies among labels. In addition, label semantic vectors are integrated to enhance label representation and enable joint modeling of text and labels. Experiments on a self-constructed Jingdezhen tourism review dataset show that the model outperforms existing methods in Micro-F1 and Hamming Loss. To further verify generalization, additional tests were conducted on the CAIL2019 Marriage and Family Elements dataset, confirming the model's effectiveness and robustness. Moreover, three sets of ablation experiments on the Jingdezhen dataset were designed to examine the contribution of the label structure and semantic fusion module, and the results demonstrated the effectiveness of these components. Overall, the findings provide useful insights for intelligent review analysis and service optimization, supporting the digital and intelligent development of the tourism industry.

Keywords: ERNIE; GCN; TextCNN; Multi-Label Text Classification; Tourism Review Analysis; Intelligent Tourism.

1. Introduction

The multi-label classification problem refers to the scenario in which a single sample may correspond to multiple category labels, and this approach has been widely applied in various domains such as text and image classification[1]. With the rapid development of deep learning, novel models have been introduced into multi-label classification tasks. For example, the TextCNN[2] model employs multi-scale convolutional kernels to capture n-gram structures in text, thereby enhancing the model's ability to understand local semantics. GCN[3] can effectively model structural relationships among labels by propagating information through graph structures, which helps to uncover potential label dependencies. Typical strategies include constructing a label co-occurrence graph to achieve semantic clustering[4] or leveraging graph attention mechanisms to strengthen inter-label feature transmission[5]. Meanwhile, pre-trained language models such as BERT[6], RoBERTa[7], and ERNIE[8], through large-scale unsupervised pre-training, have significantly improved text semantic modeling capabilities, providing a stronger semantic representation foundation for multi-label classification tasks.

In the tourism review scenario, the multi-label classification problem is particularly prominent. On mainstream tourism platforms such as Ctrip and Tongcheng, user reviews are one of the core sources of information, reflecting visitors' perceptions of scenic landscapes, dining services, accommodation experiences, and other dimensions. These reviews exhibit characteristics of semantic complexity, thematic diversity, and frequent label co-occurrence. Afzaal et al.[9] proposed a multi-label classification framework based on multi-aspect opinion mining for explicit and implicit aspect identification in tourism reviews, which demonstrated superior performance in multi-label classification tasks. Cheng and Chen[10] applied multi-label deep learning methods to identify cultural perception knowledge of Jingdezhen's heritage landscapes, highlighting the potential of multi-label classification in both tourism review analysis and cultural landscape studies. As illustrated in **Figure 1**, a single review may simultaneously correspond to labels such as "scenery," "architecture," and "catering services," which represents a typical multi-label text classification case. However, due to the flexible and diverse expressions of review texts, existing models often face challenges in capturing fine-grained semantics and contextual relations. In addition, strong correlations and co-occurrence patterns exist among labels—for example, "catering services" often co-occurs with "accommodation," while the correlation between "transportation" and "history and culture" is relatively weaker. If such label dependencies are ignored, classification models are prone to prediction bias. Moreover, tourism review data typically suffer from highly imbalanced label distributions, where certain labels have very few samples, further increasing the difficulty of model learning.

To address these challenges, researchers have recently proposed various improved approaches; however, the modeling of label semantics and structural dependencies among labels remains insufficient. Wang et al.[11] introduced a label embedding mechanism to enhance label representation, but their method did not incorporate contextual interaction modeling. Xiong et al.[12] integrated label information into the encoding structure, yet they failed to leverage inter-label dependencies or co-occurrence structures. Lu[13] combined BERT with TextCNN to capture multi-level semantics, but neglected the integration of label semantics. Yu et al.[14] developed the TLIFC-RoBERTa model, which employs a dual-tower architecture to achieve semantic fusion of texts and labels, and further utilizes an attention mechanism for interactive fusion. Nevertheless, their model does not incorporate local structural modeling methods such as TextCNN, leaving room for improvement in capturing fine-

grained textual features.

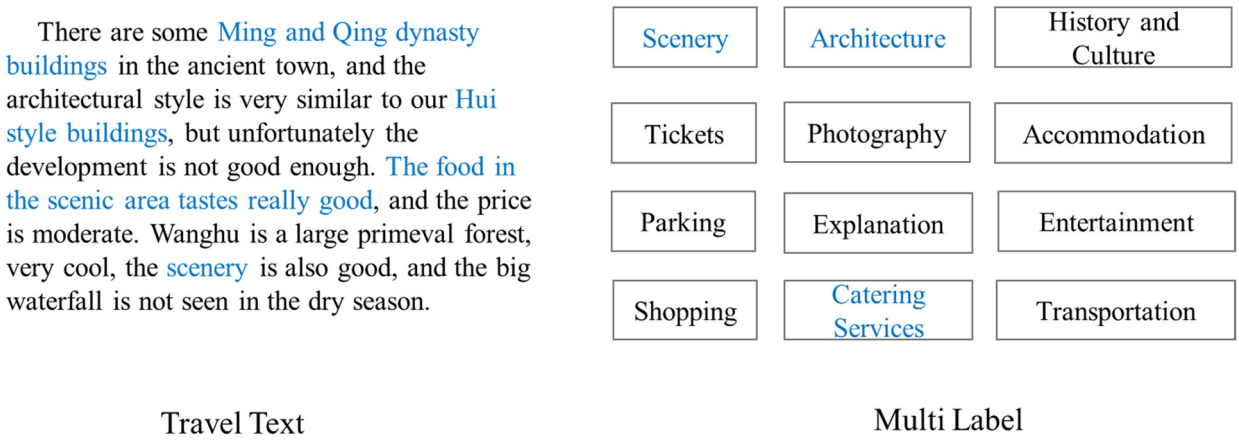

Figure 1. Multi-label recognition task for tourism texts

The main contributions of this study are as follows:

(1)We propose a multi-label text classification model, G-T-ERNIE, which integrates label information. The model leverages ERNIE to capture contextual text representations, employs TextCNN to extract local n-gram features, and enhances label representation through the joint modeling of a label co-occurrence graph and label semantic vectors. This design effectively improves the model's discriminative ability in multi-label classification tasks.

(2)We construct and annotate a Jingdezhen tourism review dataset, covering 12 categories of tourism service labels such as scenery, tickets, transportation, and explanation. This dataset provides a reliable experimental foundation for multi-label modeling in the tourism review domain.

(3) We conduct comparative experiments on both the Jingdezhen tourism review dataset and the CAIL2019 marriage and family elements dataset, systematically comparing the proposed approach with various deep learning methods. Experimental results show that the proposed model significantly outperforms existing methods in terms of Micro-F1, Hamming Loss, and other metrics, demonstrating not only superior performance in tourism review scenarios but also strong cross-domain generalization ability and robustness in legal text classification tasks.

2. Related Work

2.1. ERNIE Pre-Trained Language Model

ERNIE enhances Chinese semantic understanding by introducing knowledge masking and entity-level training tasks. Compared with BERT, ERNIE 3.0[15] achieves superior performance on the SuperGLUE benchmark and, in short-text classification tasks, effectively captures the semantic associations between labels and texts through dynamic word embeddings. Recent studies have applied ERNIE in domains such as biomedical research[16] and sentiment analysis[17], further validating its cross-domain adaptability.

As a pre-trained model, ERNIE 3.0 processes tokenized and segmented text sequences by combining word embeddings, positional embeddings, and segment embeddings to generate context-aware semantic representations[18]. The architecture of ERNIE 3.0 is illustrated in **Figure 2**.

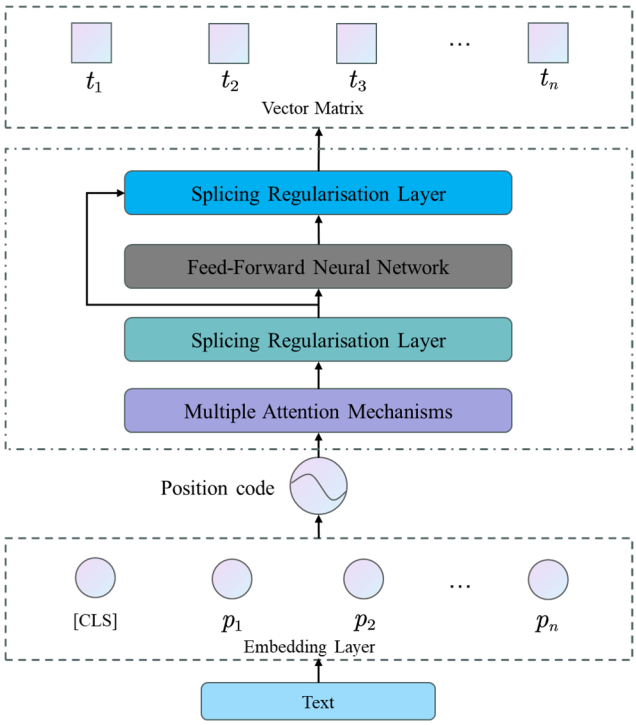

Figure 2. The architecture of the ERNIE 3.0 model

2.2. Research on the Integration of Pre-trained Models and TextCNN

The TextCNN model, proposed by Kim in 2014, employs convolutional kernels of different sizes to process input texts, enabling the extraction of n-gram semantic features at multiple scales. A max-pooling operation is then applied to select the most representative local information. These features are concatenated and passed into fully connected layers for classification tasks, thereby improving the model's ability to capture local semantic structures. Owing to its simplicity, low parameter count, and strong performance in multi-class text classification scenarios, TextCNN is often adopted as an effective complementary module to pre-trained language models. For instance, Zheng et al.[19] employed the ALBERT pre-trained language model to generate dynamic character embeddings and integrated TextCNN to capture semantic features at different levels of abstraction for multi-label medical text classification. Yan[20] proposed the ERNIE-TextCNN model, which combines the contextual

semantic representations of the ERNIE pre-trained model with the convolutional feature extraction capability of TextCNN to enhance the performance of Chinese news headline classification.

2.3. GCN Model

GCN is a deep learning model specifically designed to process graph-structured data and is adept at uncovering latent relationships among nodes. In multi-label text classification, labels often exhibit complex dependencies and correlations. By constructing a label co-occurrence graph, GCN treats each label as a node and assigns edge weights based on co-occurrence frequency. Graph convolution operations are then applied over this structure to enable information interaction and propagation among label semantics, thereby generating label representation vectors with stronger structural expressiveness. For example, Meng et al.[21] proposed the MFLSCI model for multi-label legal text classification, which incorporates GCN to learn node representations by aggregating information from neighboring nodes. By combining multi-granularity textual information with label semantic correlations, their approach effectively captured both textual semantics and inter-label dependencies, leading to improved classification performance.

3. Multi-Label Text Classification Model Based on G-T-ERNIE

3.1. Overall Model Framework

The proposed G-T-ERNIE multi-label text classification model is designed to integrate contextual semantic features, n-gram local features, structural co-occurrence information among labels, and the semantic representations of labels themselves, thereby improving the accuracy of multi-label prediction and the discriminative capability among labels. As illustrated in **Figure 3**, the overall architecture of the model consists of three core modules: the textual comprehensive semantic representation module, the label comprehensive semantic representation module, and the text–label matching and multi-label classification module.

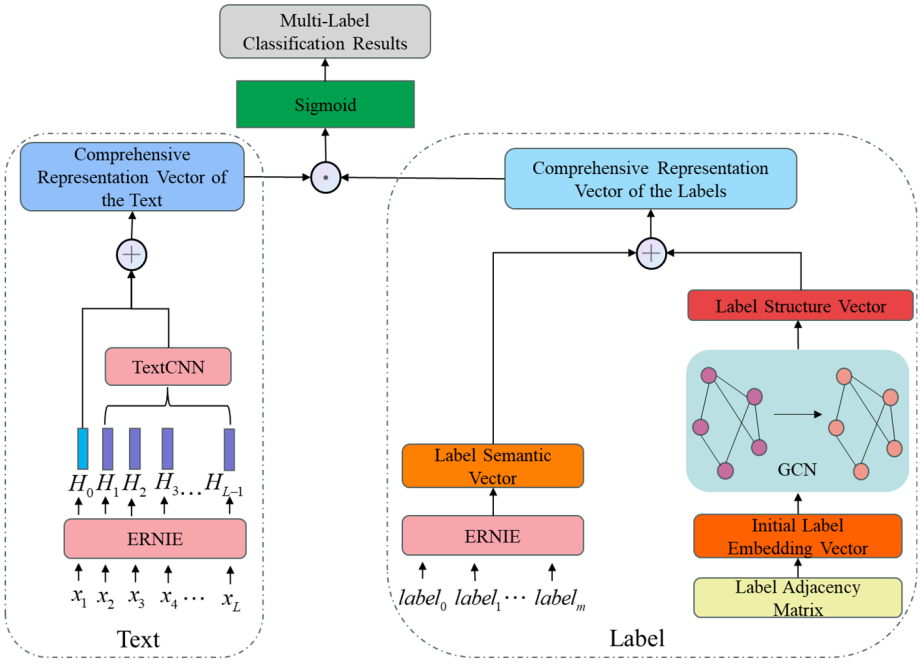

Figure 3. The architecture of the G-T-ERNIE model

For the textual part, the input reviews are first tokenized and indexed, then passed into the pre-trained language model ERNIE to obtain contextual semantic representations. The first position vector (i.e., [CLS]) is used to represent the global semantics of the review. To further enhance the modeling of local information, a TextCNN module is introduced. The TextCNN receives the token sequence excluding the [CLS] token, applies multiple convolutional kernels of different sizes (e.g., 2, 3, 4) to extract n-gram features, and employs max pooling to obtain the most salient local features. The outputs of TextCNN are concatenated with the [CLS] vector to form a comprehensive textual representation.

For the label part, fixed label names such as “scenery,” “tickets,” and “shopping” are treated as short text inputs to the ERNIE model, from which their [CLS] semantic vectors are extracted. After linear projection, these vectors serve as the semantic representations of the labels. Meanwhile, the co-occurrence relationships among labels are calculated from the training set to construct a label adjacency matrix, which is then processed using a two-layer GCN to propagate structural information among labels. The structural vectors output by GCN are aligned in dimension and concatenated with the semantic vectors, then fused via a multilayer perceptron to generate the final comprehensive label representations.

Finally, the textual and label vectors obtained in the same embedding space are matched by computing their dot-product similarity scores. These scores are normalized through a Sigmoid function to yield the final prediction probabilities for multi-label classification.

3.2. Textual Comprehensive Semantic Representation Module: Integration of ERNIE and TextCNN

To fully exploit both the global semantic information and

the local key features contained in user reviews, this study employs an "ERNIE + TextCNN" architecture for semantic encoding of the input text. The framework is illustrated in **Figure 4**.

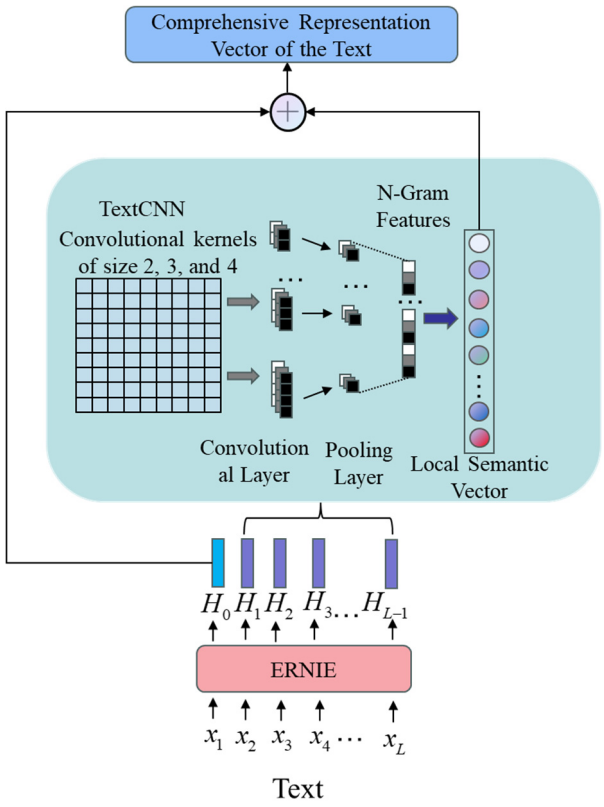

Figure 4. Framework of the textual comprehensive semantic representation module

First, the user review text is tokenized and index-encoded to obtain the input sequence $X = \{x_1, x_2, \ldots, x_L\}$ .The sequence is then fed into the pre-trained language model ERNIE to generate the contextual semantic output representations:

$$H = \text{ERNIE}(X) \tag{1}$$

Among them, H_0, i.e., the vector of the first token [CLS], denoted as h_{cls} , is regarded as the global semantic representation of the review.

On this basis, TextCNN is further employed to capture local short-text patterns within the review. For the remaining token sequence $H_{1:L-1}$, multiple convolutional kernels $\{f_i\}_{i=1}^{k}$ with different sizes are applied, followed by one-dimensional convolution and max-pooling operations, to extract feature vectors p_i at different granularities:

$$p_i = \max\left(\text{ReLU}(\text{Conv}_{f_i}(H_{1:L-1}))\right) \tag{2}$$

Finally, the [CLS] vector h_{cls} is concatenated with the outputs of all convolutional kernels to form the fused semantic representation of the text:

$$t = [h_{cls}; p_1; \ldots; p_k] \tag{3}$$

This module not only preserves the contextual dependencies of long texts but also enhances the representation of local structures, thereby providing a solid semantic foundation for subsequent label matching.

3.3. Label Comprehensive Semantic Representation Module: Fusion of GCN and Label Semantic Vectors

To fully capture both the structural relationships and semantic characteristics of labels, this study integrates semantic vectors of label names encoded by the ERNIE model with the structural representations propagated through the GCN.

First, a label co-occurrence adjacency matrix A is constructed based on the 12 categories of labels in the training set, and the embedding vector of each label is initialized as $E^{(0)}$. Then, a two-layer GCN is employed to model the structural relationships among labels, with the propagation process defined as follows:

$$E^{(l+1)} = \sigma\left(\hat{D}^{-1/2}\hat{A}\hat{D}^{-1/2}E^{(l)}W^{(l)}\right) \tag{4}$$

where $\hat{A} = A + I$ denotes the adjacency matrix with self-loops, $\hat{D}$ is the corresponding degree matrix, and $W^{(l)}$ represents the trainable parameters. After the propagation process, the structural representation of the labels is obtained as $E^{(2)}$.

Meanwhile, labels (e.g., "scenery", "tickets") are treated as independent short texts and fed into the ERNIE model. The token representations produced by ERNIE are aggregated through masked average pooling to obtain the semantic representation of each label s_i . A linear transformation W_s is then applied to project these representations into the label embedding space with reduced dimensionality.

$$s_{i'} = W_s \cdot s_i \tag{5}$$

The structural representation and semantic representation are concatenated and fed into the MLP for feature fusion.

$$z_i = \text{MLP}\left(\left[E_i^{(2)}; s_{i'}\right]\right) \tag{6}$$

This fusion strategy enables the model to capture both the structural characteristics of label co-occurrence and the semantic information embedded in the labels. For example, in the Jingdezhen tourism review dataset, the label "history and culture" often co-occurs with labels such as "architecture" and "explanation", which can be modeled through the label co-occurrence graph. At the same time, the label "history and culture" can obtain semantic embeddings from ERNIE, thereby forming a more comprehensive representation of the label.

3.4. Text–Label Matching and Multi-Label Classification Module

The comprehensive representation vector of the text is matched with that of the label through dot product, and the resulting score is used to measure the relevance between the text and the label.

$$\hat{y}_i = \sigma(t^{\cdot} z_i), \quad \text{for } i = 1, 2, \ldots, C \tag{7}$$

Here, $\sigma(\cdot)$ denotes the Sigmoid function, and $\hat{y}_i \in (0,1)$ represents the matching probability between the text and the i -th label. Essentially, this approach follows a dualtower matching strategy, where the comprehensive representation vector of the text is compared with that of the label within a shared embedding space. This design is wellsuited to the

"one-to-many" modeling requirements of multi-label classification tasks.

During the inference phase, the model performs a binarization operation based on the predicted probability $\hat{y}_i$. A global fixed threshold is applied to determine whether a label is relevant, and the calculation is defined as follows:

$$\overline{y}_i = \begin{cases} 1, & \hat{y}_i \geq \tau, \\ 0, & \hat{y}_i < \tau, \end{cases} \quad \text{for } i = 1, 2, \ldots, C, \quad \tau = 0.5 \quad (8)$$

Specifically, when the predicted probability of a label is greater than or equal to 0.5, it is classified as a positive sample; otherwise, it is considered a negative sample. This strategy ensures the interpretability and consistency of the prediction results, while avoiding the recall degradation caused by an excessively high threshold and the false detections resulting from an overly low threshold.

To address the issue of label imbalance in practical applications, a weighted binary cross-entropy loss function is introduced during training:

$$\mathrm{L} = -\sum_{i=1}^{C} \omega_i [y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i)] \quad (9)$$

Here, ω_i denotes the loss coefficient for label i, which is adjusted according to the sample frequency to prevent high-frequency labels from dominating the training process. For instance, in the Jingdezhen tourism review dataset, labels such as "photography" and "scenery" occur with relatively high frequency, whereas low-frequency labels (e.g., "parking") would struggle to learn effective representations without weighting.

4. Experiments

4.1. Description of Datasets

Considering the limited sample size of the Jingdezhen tourist review dataset and the current lack of other Chinese multi-label datasets for tourist reviews, we also selected the marriage and family elements dataset from CAIL2019[22] to verify the generalization performance of the multi-label classification model in this paper. The specific details of the dataset are as follows:

(1) Jingdezhen tourist review dataset: A total of 16,000 pieces of comment texts about Jingdezhen scenic spots were crawled from Ctrip and Tongcheng Travel websites using Python crawlers as the original experimental dataset. After data cleaning operations such as removing duplicate comments and invalid content, data preprocessing like compressing repeated characters or words was performed. Finally, 8280 pieces of data were obtained and manually annotated[23]. Based on the sample content, 12 label categories were set in this paper. The number of samples under each category is shown in **Figure 5**.

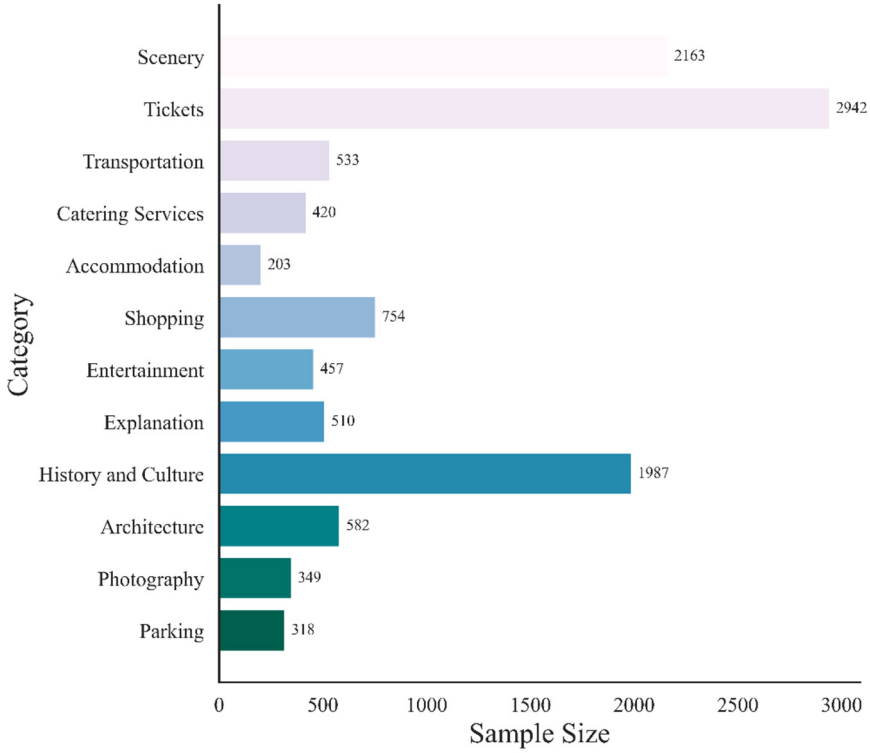

Figure 5. Distribution of sample size by category

(2) CAIL2019 Marriage and Family Elements Dataset[22]: This dataset is derived from public legal documents on the "China Judgments Online". Each sentence in the dataset is labeled with corresponding category tags, including 20 categories such as "payment of child support", "having joint marital debts", "statutory divorce", and "post-marital personal property".

The Jingdezhen tourist review dataset is divided into training, validation, and test sets in an 8:1:1 ratio. The specific division and label counts of the two datasets are shown in Table 1.

4.2. Evaluation Metrics

The following metrics are employed in this paper to evaluate the model performance:

$$\text{Micro-F1} = \frac{2 \cdot \text{Micro-Precision} \cdot \text{Micro-Recall}}{\text{Micro-Precision} + \text{Micro-Recall}} \quad (10)$$

Table 1. Experimental dataset division and label counts

Datasets	Training	Validation	Test	Number of Labels
Jingdezhen Tourist Review Dataset	6624	828	828	12
CAIL2019 Marriage and Family Elements Dataset	12766	1611	1611	20

$$\text{Hammingloss} = \frac{1}{M} \cdot \sum_{i=1}^{M} \frac{1}{q} | P_{i\Delta} Y_i | \quad (11)$$

In evaluation metrics, Micro-F1 is often used to assess datasets with uneven category distributions. It takes into account the overall precision and recall of all labels. Hamming loss[24] is used to evaluate the proportion of misclassified labels in the instances. Here, q represents the number of labels corresponding to a document, and P_i and Y_i denote the predicted label set and the true label set of the corresponding document, respectively. Δ represents the symmetric difference between them.

4.3. Comparative Experiment Setup

To verify the effectiveness of the G-T-ERNIE model in multi-label classification tasks, the following comparative models are selected:

(1) ERNIE: The Chinese pre-trained language model ERNIE 3.0 released by Baidu is used to encode input texts, extract the global semantic representation at the [CLS] position, and finally connect a fully connected layer for multi-label classification, serving as a benchmark method for pure semantic modeling.

(2) ALBERT: A lightweight pre-trained model ALBERT is adopted, which significantly reduces the model parameter scale while maintaining performance through a parameter sharing strategy, thereby improving training efficiency. It encodes input texts, extracts the semantic representation at the [CLS] position, and feeds it into a fully connected layer for label prediction.

(3) TextCNN: Static Word2Vec word vectors trained on Sogou News corpus are used for encoding. Multiple convolution kernels of different sizes are employed to extract local semantic information in the text. Then, max-pooling is used to select the most important features from each channel, which are concatenated and fed into a fully connected layer for multi-label classification.

(4) BiLSTM: It also uses static Word2Vec word vectors trained on Sogou News corpus for encoding. The difference is that it employs a bidirectional LSTM network, a structure that can better capture the associations between contexts, especially suitable for processing sequential data like text. Finally, a fully connected layer is added for multi-label prediction.

(5) BERT-BiLSTM: The Chinese pre-trained language model BERT is used as an encoder to extract contextual representations of the text, and its output is fed into a bidirectional LSTM network to further model temporal dependencies. The output of BiLSTM is then sent to a fully connected layer to realize label prediction.

(6) ALBERT-TextCNN: The lightweight pre-trained model ALBERT is used as an encoder to obtain contextual semantic representations of the text, and then the TextCNN structure is used to extract local features, achieving the fusion of semantic information at different granularities.

(7) TLIFC-RoBERTa: The Chinese pre-trained language model chinese-roberta-wwm-ext provided by the Harbin Institute of Technology and iFlytek Joint Laboratory is adopted. The WWM (Whole Word Masking) strategy refers to masking the entire part after tokenizing the text[25]. The text and label-trained data are input into the Attention layer and the label extraction module, and finally, adaptive fusion classification is performed.

4.4. Experimental Environment and Parameters

The experiments were conducted on an NVIDIA GeForce RTX 4090 GPU to provide sufficient computing resources for model training and evaluation. The development environment was based on Anaconda3 (Python 3.8.19), with PyTorch 2.2.2 used as the deep learning framework.

For text encoding, Baidu's Chinese pre-trained language model ERNIE-3.0-Base was adopted, featuring a hidden layer dimension of 768 and a maximum text length of 512. The TextCNN module utilized convolution kernels of sizes 2, 3, and 4, with 128 channels for each configuration to extract local features at different scales. The ReLU activation function was employed, and a dropout rate of 0.5 was set to mitigate overfitting.

Considering that the parameters of the pre-trained language model ERNIE have converged through large-scale corpus training, while other modules (such as TextCNN, GCN, and the fusion layer) were newly initialized with insufficiently trained parameters, a layered learning rate strategy was adopted. The ERNIE module used a smaller learning rate of 1e-5 to avoid damaging its existing semantic representation capabilities, while the remaining modules adopted a relatively larger learning rate of 5e-4 to accelerate the convergence of new parameters. The Adam optimizer was used for the model.

During training, the batch size was set to 8, the number of iterations (epochs) was 10, and the threshold was 0.5. The loss function was weighted binary cross-entropy, with weight coefficients of 0.5 and 1.0 for positive and negative samples, respectively, to alleviate the impact of class imbalance.

4.5. Comparative Experiment Results and Analysis

The experimental results of the model on the test sets of the two datasets are presented in **Table 2** and **Table 3**.

In terms of the Micro-F1 metric, the G-T-ERNIE model achieved the highest value of 0.890 on the Jingdezhen tourist review dataset, which is nearly 2 percentage points higher than the second-ranked TLIFC-RoBERTa model. This indicates that the proposed model has obvious advantages in overall classification accuracy. Compared with models built using traditional Word2Vec word vectors (such as TextCNN and BiLSTM), G-T-ERNIE also shows a significant improvement, demonstrating that the introduction of pre-trained models and label structure information plays a positive role in enhancing model performance.

Among the comparative models, the Micro-F1 of the ALBERT model is 0.808, which is significantly lower than that of the ERNIE model (0.866). Although ALBERT is more lightweight in terms of parameter quantity, its semantic modeling ability is slightly inferior to ERNIE, possibly

because its parameter sharing mechanism reduces the ability to express complex semantic relationships.

Table 2. Comparison of model performance on Jingdezhen tourist review dataset

Models	Micro-Precision	Micro-Recall	Micro-F1	Hammingloss
ERNIE	0.819	0.919	0.866	0.0348
ALBERT	0.768	0.853	0.808	0.0495
TextCNN	0.866	0.859	0.862	0.0335
BiLSTM	0.798	0.627	0.702	0.0650
BERT-BiLSTM	0.826	0.919	0.870	0.0335
ALBERT-TextCNN	0.778	0.849	0.812	0.0481
TLIFC-RoBERTa	0.820	0.928	0.871	0.0337
G-T-ERNIE	0.847	0.937	0.890	0.0284

The ALBERT-TextCNN model introduced a local convolution feature extraction mechanism on the basis of ALBERT, and its Micro-F1 increased to 0.812. This shows that combining the pre-trained model with the convolutional neural network can alleviate the insufficient expressive ability of ALBERT to a certain extent and enhance the ability to capture local semantics of the text.

The BERT-BiLSTM model performed balanced in various metrics, with a Micro-F1 of 0.870, indicating that introducing BiLSTM for temporal modeling of BERT encoding results is helpful to improve the ability to capture context. However, its Hamming loss is 0.0335, which is slightly higher than that of G-T-ERNIE, indicating that there are still certain deficiencies in its fine-grained label prediction.

In addition, G-T-ERNIE has the lowest Hamming loss value of only 0.0284, reflecting that the model has smaller prediction errors and stronger discriminative ability.

Table 3. Comparison of model performance on CAIL2019 marriage and family elements dataset

Models	Micro-Precision	Micro-Recall	Micro-F1	Hammingloss
ERNIE	0.855	0.912	0.883	0.0216
ALBERT	0.820	0.891	0.854	0.0272
TextCNN	0.893	0.877	0.885	0.0204
BiLSTM	0.819	0.820	0.819	0.0322
BERT-BiLSTM	0.868	0.929	0.897	0.0190
ALBERT-TextCNN	0.862	0.882	0.872	0.0231
TLIFC-RoBERTa	0.893	0.914	0.903	0.0175
G-T-ERNIE	0.885	0.930	0.907	0.0171

On the CAIL2019 Marriage and Family Elements Dataset, the advantages of the G-T-ERNIE model remain evident. G-T-ERNIE outperforms other comparative models in both core metrics of Micro-F1 and Hamming loss, especially demonstrating stronger robustness and generalization ability in scenarios with uneven label distribution or strong correlations between labels. Its overall performance is superior to current mainstream methods, indicating that fusing label semantic and structural information helps enhance the model's ability to model complex label relationships, fully verifying the effectiveness of the G-T-ERNIE model in multi-label text classification tasks.

4.6. Ablation Experiment Results and Analysis

To further analyze the actual contribution of the label structure and semantic fusion module in the G-T-ERNIE model, three sets of ablation experiments were designed and conducted on the Jingdezhen tourist review dataset. The results of the ablation experiments are shown in **Table 4**:

(1) G-T-ERNIE(-GCN-Label): The label structure and semantic fusion module of the proposed model are removed, retaining only ERNIE and TextCNN.

(2) G-T-ERNIE(-GCN): The GCN part of the label structure and semantic fusion module in the proposed model is removed, while the label semantic representation is retained.

(3) G-T-ERNIE(-Label): The label semantic representation part of the label structure and semantic fusion module in the proposed model is removed, while GCN for modeling label structure is retained.

Table 4. Ablation experiment results on Jingdezhen tourist review dataset

Models	Micro-Precision	Micro-Recall	Micro-F1	Hammingloss
G-T-ERNIE(-GCN-Label)	0.837	0.928	0.880	0.0309
G-T-ERNIE(-GCN)	0.816	0.943	0.875	0.0329
G-T-ERNIE(-Label)	0.856	0.921	0.887	0.0287
G-T-ERNIE	0.847	0.937	0.890	0.0284

The results of the ablation experiments show that the G-T-ERNIE model achieves optimal performance in all metrics after introducing label semantic information and label structure modeling.

When both label structure and label semantic information are removed, leaving only the basic model composed of ERNIE and TextCNN, the overall performance decreases but still outperforms the ERNIE model in the comparative experiments. This indicates that the combination of pre-trained models and convolutional neural networks has certain per-formance advantages in multi-label tasks.

The model performance declines significantly after removing label structure information, which suggests that the co-occurrence relationships between labels play an important role in multi-label classification tasks.

When label structure is retained but label semantic representation is removed, the model accuracy decreases slightly, indicating that semantic information helps enhance the model's discriminative ability.

Overall, the fusion of label semantic and structural information plays a key role in improving the model's discriminative ability and generalization performance.

5. Conclusion

For the multi-label classification task of tourist review texts, this paper proposes a G-T-ERNIE model that integrates label structure and semantic information. Based on the global

semantic representation provided by the ERNIE pre-trained language model, the model combines the TextCNN module to extract local semantic information. Meanwhile, it introduces GCN to capture the structural dependencies between labels, fuses label semantic information, and enhances the ability of semantic expression.

Experimental results on the Jingdezhen tourist review dataset and the CAIL2019 Marriage and Family Elements Dataset show that the G-T-ERNIE model outperforms baseline methods in multiple evaluation metrics, demonstrating excellent multi-label text classification capabilities. The model exhibits strong adaptability in handling complex classification tasks with a large number of labels and obvious semantic overlaps, thus possessing certain practical application potential. In future work, we will further optimize the model structure to adapt to longer review texts and explore the introduction of an attention mechanism to improve the model's ability to model long texts.

Acknowledgments

This work was supported in part by a grant from the National Natural Science Foundation of China (No. 62162032), China Ceramic Development Research Institute of Jingdezhen Ceramic University, Jiangxi Social Science Planning Commission Project(2023ZK03).

References

[1] Dongmei, L.; Yu, Y.; Xianghao, M.; Xiaoping, Z.; Chao, S.; Yufeng, Z. Review on multi-lable classification. Journal of Frontiers of Computer Science & Technology 2023, 17, 2529.

[2] Kim, Y. Convolutional neural networks for sentence classification. arXiv 2014, arXiv:1408.5882.

[3] Kipf, T. Semi-Supervised Classification with Graph Convolutional Networks. arXiv 2016, arXiv:1609.02907.

[4] Zong, D.; Sun, S. GNN-XML: graph neural networks for extreme multi-label text classification. arXiv 2020, arXiv: 2012. 05860.

[5] Pal, A.; Selvakumar, M.; Sankarasubbu, M. Multi-label text classification using attention-based graph neural network. arXiv 2020, arXiv:2003.11644.

[6] Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers), 2019; pp. 4171-4186.

[7] Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; Stoyanov, V. Roberta: A robustly optimized bert pretraining approach. arXiv 2019, arXiv:1907.11692.

[8] Zhang, Z.; Han, X.; Liu, Z.; Jiang, X.; Sun, M.; Liu, Q. ERNIE: Enhanced language representation with informative entities. arXiv 2019, arXiv:1905.07129.

[9] Afzaal, M.; Usman, M.; Fong, A.C.; Fong, S. Multiaspect-based opinion classification model for tourist reviews. Expert Systems 2019, 36, e12371.

[10] Cheng, Y.; Chen, W. Cultural Perception of Tourism Heritage Landscapes via Multi-Label Deep Learning: A Study of Jingdezhen, the Porcelain Capital. Land 2025, 14, 559.

[11] Wang, G.; Li, C.; Wang, W.; Zhang, Y.; Shen, D.; Zhang, X.; Henao, R.; Carin, L. Joint embedding of words and labels for text classification. arXiv 2018, arXiv:1805.04174.

[12] Xiong, Y.; Feng, Y.; Wu, H.; Kamigaito, H.; Okumura, M. Fusing label embedding into bert: An efficient improvement for text classification. In Proceedings of the Findings of the association for computational linguistics: ACL-IJCNLP 2021, 2021; pp. 1743-1750.

[13] Lu J. Multi-label classification method of open source threat intelligence text based on Bert-TexCNN. Journal of Information Security Research 2024, 10, 760-768; http://www. sicris. cn/CN/Y2024/V10/I8/760.

[14] Yu H; Zhou Y; Zhai M; Liu H. Text classification based on pre-training model and label fusion. Journal of Computer Applications 2024, 44, 709-714; https://www.joca. cn/ CN/ 10. 11772/j.issn.1001-9081.2023030340.

[15] Sun, Y.; Wang, S.; Feng, S.; Ding, S.; Pang, C.; Shang, J.; Liu, J.; Chen, X.; Zhao, Y.; Lu, Y. Ernie 3.0: Large-scale knowledge enhanced pre-training for language understanding and generation. arXiv 2021, arXiv:2107.02137.

[16] Li, Z.; Ren, J. Fine-tuning ERNIE for chest abnormal imaging signs extraction. Journal of Biomedical Informatics 2020, 108, 103492.

[17] Sun, Y.; Yu, Z.; Sun, Y.; Xu, Y.; Song, B. A novel approach for multiclass sentiment analysis on Chinese social media with ERNIE-MCBMA. Scientific Reports 2025, 15, 18675.

[18] Li, B.; Hou, Y.; Dong, J.; Yang, B.; Wang, X. ICRA: A study of highly accurate course recommendation models incorporating false review filtering and ERNIE 3.0. PloS one 2024, 19, e0313928.

[19] Zheng, C.; Wang, X.; Wang, T.; Deng, Y.; Yin, T. Multi-label classification for medical text based on ALBERT-TextCNN model. Journal of Shandong University (Natural Science) 2022, 57, 21-29.

[20] Yan, Y. ERNIE-TextCNN: research on classification methods of Chinese news headlines in different situations. Scientific Reports 2025, 15, 29071.

[21] Meng, C.; Todo, Y.; Tang, C.; Luan, L.; Tang, Z. MFLSCI: Multi-granularity fusion and label semantic correlation information for multi-label legal text classification. Engineering Applications of Artificial Intelligence 2025, 139, 109604.

[22] Xiao, C.; Zhong, H.; Guo, Z.; Tu, C.; Liu, Z.; Sun, M.; Zhang, T.; Han, X.; Hu, Z.; Wang, H. CAIL2019-SCM: a dataset of similar case matching in legal domain. arXiv 2019, arXiv:1911.08962.

[23] Hsieh, Y.-H.; Zeng, X.-P. Sentiment analysis: An ERNIE-BiLSTM approach to bullet screen comments. Sensors 2022, 22, 5223.

[24] Cheng, Q.; Shi, W. Hierarchical multi-label text classification of tourism resources using a label-aware dual graph attention network. Information Processing & Management 2025, 62, 103952.

[25] Cui, Y.; Che, W.; Liu, T.; Qin, B.; Wang, S.; Hu, G. Revisiting pre-trained models for Chinese natural language processing. arXiv 2020, arXiv:2004.13922.

[26] Peng, Y.; Wu, W.; Ren, J.; Yu, X. Novel GCN model using dense connection and attention mechanism for text classification. Neural Processing Letters 2024, 56, 144.