Computer Life
ISSN: 1819-4818 | Vol. 12, No. 1, 2024

Bert-based model for Ceramic Emotional Text Recognition and Multi-label Classification

Busheng Li, Siqi Li, Wangren Qiu*, Tao Liu

School of Information Engineering, Jingdezhen Ceramic University, Jingdezhen, 333000, China
* **Corresponding author**: Wangren Qiu (Email: qiuone@163.com)

Abstract: Ceramic products are necessities for daily home life, while the emotional analysis research on ceramic products cannot keep up with the needs of social and economic development. To mine the valuable information on ceramic emotion quickly and effectively, this work has proposed a novel method for text recognition and classification of microblogging information related to Jingdezhen ceramics by fusing Bert model and multi-label classifiers on the novel dataset we first established. Firstly, the first multi-label emotion dataset consisting of 7564 samples was constructed on 10154 raw samples obtained with Python from microblogging information, which is noted as the keyword “Jingdezhen ceramics”. Secondly, the data were categorized with multiple labels into useful or useless information. Useful information was further classified into three categories: price information, appearance information, and quality information. Thirdly, a hybrid model combining Bert and multi-label classifiers has been provided for analyzing ceramic emotion. Based on a series of experiments, it has been observed that the predictor’s accuracy is 92% and its loss value is 0.33. This work highlights the annotation guidelines, dataset characteristics, and insights into different methodologies used for analyzing ceramic emotion and cultural heritage.

Keywords: BERT model, multi-label classification, social media, sentiment analysis, deep learning

1. Introduction

Microblog is a microblogging platform that is used by several hundred million people daily to express themselves, share opinions, or store information; and due to its wide usage, it is recognized as an ideal platform for studying emotions or predicting the outcomes of experimental interventions [1]. Studying emotions in the text could help us to understand the behavior of individuals and give us the key to the understanding of people's feelings and perceptions. Using text recognition technology, massive amounts of text data can be processed and used rapidly and accurately, which will help the information society change and advance.

In recent years, the technology of text mining has been extensively used in the fields of artificial intelligence, data mining, and information retrieval. However, it is rarely utilized in the field of ceramics. In 2010, Yu applied this technique to the ceramic industry and by text categorization he realized the ceramic information document classification and efficient reading of the information retrieved [4]. In 2013, Zhao analyzed the expression of emotion elements in the design of domestic ceramics based on the successful cases at home and abroad [5]. In 2023, Yang advanced a classification model of the ancient ceramic decoration, which improved the accuracy of ancient ceramic decoration classification [6]. In 2023, Yan proposed five data mining methods based on a study of the chemical composition of ancient ceramics and ceramic raw materials, and whereby carried out application research on ancient ceramic dating and ceramic raw material classification [7]. Although many researchers have conducted further studies on the basis of the previous work, there still exist some problems need to be solved: (1) The scope of ceramic text research is quite limited, and the information comes largely from a single source, with the majority of it being from the cost-free ceramic-related literature and ceramic works. (2) Text classification is not so duly valued as text translation or text manipulation in ceramic text research. (3) Research on the multi-label classification of ceramic texts is seriously insufficient; and though there are multiple ceramic categories concerned in the existing research, but only one category can be labeled for the sample.

Currently, text classification research focuses on the task of binary or triple classification of social media information, with each piece of data covering only one type of information. In actuality, social media messages may deal with multiple topics at once, which means that a single piece of message may contain multiple categories of information, which is more in line with the characteristics of actual social media messages. Multi-label classification allows multiple categories of samples to be labeled at the same time, which is conducive to capturing diverse information content more accurately and improving the granularity and accuracy of classification [8]. The conventional text classification approaches include supervised learning, unsupervised learning, and semi-supervised learning techniques [10]. These models are mainly concerned with probability or information content and their performance depends on the size of the corpus. However, these models cannot provide a good comprehension of the semantics of the text for modeling and classifying large-scale texts [12]. Due to the rapid development of information networks and advances in computing power, deep learning-based techniques are already commonly employed in text classification tasks and are gradually becoming an important research area for academics [13]. The BERT (Bidirectional Encoder Representations from Transformers) pre-training model was introduced by Google in 2018. It uses a significant amount of data and model parameters for unsupervised pre-training and supervised fine-tuning based on a particular language-related task. The fine-tuned model can capture multiple meanings of words, grammatical structures, sentence structures, realistic connotations, and other aspects of the text. This deep learning approach based on "pre-training + fine-tuning" has become a hot research topic.

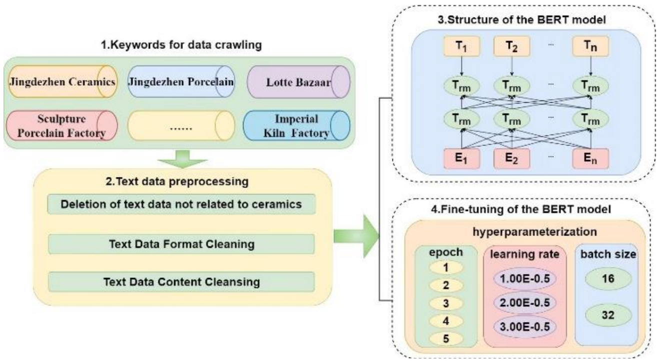

Figure 1. Framework of BERT based and multi-label classification model

In this study, text data related to ceramics, particularly Jingdezhen ceramics, posted by microblogging users are collected as training samples. The BERT pre-training model is used to create a multi-hashtag classification model, which is intended to swiftly extract precise data regarding ceramic perceptions in social media [16]. The study's findings have promoted the standard improving and cultural significance of ceramic goods while closing the gap in the field of text emotion recognition in ceramics [17]. The BERT multi-label classification model is illustrated in Fig. 1.

2. Research Methodology

2.1. Data crawling

The dataset was crawled by Python with the keywords " Jingdezhen ceramics" and " Jingdezhen porcelain" from microblog. Table 1 shows the profile of data crawling. The text data gathered consists of the microblog ID, the publisher's moniker, the microblog's content, and the moment the four fields of data were released. These fields can be used to classify and analyze social media information to help understand and mine content and features of the text [18].

Table 1. The profile of the data set

Keywords	Number of texts	Keywords	Number of texts
Jingdezhen Ceramics	697	Lotte Bazaar	936
Jingdezhen Porcelain	1173	Tao Xichuan	718
Jingdezhen Ancient Porcelain	574	Imperial Kiln Museum	746
Jingdezhen Ancient Kiln	535	Porcelain Palace	523
Jingdezhen China Ceramics Museum	790	Tao Ran Jie	501
Sculpture Porcelain Factory	712	Porcelain Fair	522
Imperial Kiln Porcelain Factory	788	Shamrock	448
Imperial Kiln Factory	117	Tao Yangli	109

The information obtained for the study was limited to short texts and did not include videos, images, extended texts, or other non-text messages. After data collection, we normalized the text data and removed the digits, punctuation marks, and the words for pausing at the word-and-character-level normalization. With such normalization, the text data is more consistent, readable, and analyzable. As the quality of the data will be impacted by some redundant information inevitable in the results of keyword-based data crawling, manual judgment and redundant data elimination are needed. Table 2 shows an example of text after removing extraneous data. Table 3 shows an example depiction of the raw data and a portion of the cleaned data.

Table 2. Some examples of text after removing extraneous data

Raw text data	Text data after deletion of irrelevant information
I was impressed by the beauty of the porcelain in the alley next to the Imperial Kiln Factory – Jingdezhen.	I was impressed by the beauty of the porcelain in the alley next to the Imperial Kiln Factory - Jingdezhen
Jingdezhen Ceramic University, what is the color test in today's reexamination 🙏🙏🙏🙏🙏🙏🙏🙏	
Jingdezhen ceramics, whether it is the Ming Dynasty or the Qing Dynasty, Kang Yong's porcelain bowl looks very elegant and pure.	Jingdezhen ceramics, whether it is the Ming Dynasty or the Qing Dynasty, Kang Yong's porcelain bowl looks very elegant and pure.
Jingdezhen ceramic artisans hand-painting gold, this craft is amazing #NewYearRhapsody	Jingdezhen ceramic artisans hand-painting gold, this craft is amazing #NewYearRhapsody

Table 3. Some examples of the raw data and a portion of the cleaned data

Raw text data	Cleaned text data
🔍 Today I wandered through the sculpture porcelain factory's Lotte bazaar only to realize that it's more expensive than Tao Xichuan 😂	today wandered through Lotte bazaar realize sculpture porcelain factory's more expensive Tao Xichuan
A half-day excursion to Tao Xichuan. 🍃🏺Sculpture Porcelain Factory → 🥰Love Jingdezhen's delicate and aesthetically pleasing little works of art!	half-day excursion Tao Xichuan sculpture Porcelain Factory love Jingdezhen's delicate aesthetically little works of art
Lotte Bazaar is full of artisans, the ceramics here look good in any photo but it's expensive to just pick them out 😂 #Jingdezhen #HandsOnPhoto.	Lotte Bazaar full artisans ceramics look good photo Expensive pick out Jingdezhen HandsOnPhoto

2.2. Data annotation

Multi-label classification is frequently used in text mining to assign text to numerous categories or topic labels [20]. It is important to note that this classification can be applied in several domains instead of a single domain. For instance, based on the multi-label classification method, deep learning models such as convolutional neural networks, recurrent neural networks, and attention mechanisms have produced impressive results to date.

One of the biggest difficulties, also the crucial step, in multi-label classification is mining the correlation between labels, and researchers have figured out several methods to model and capture this correlation [21]. For constructing the predicting model, techniques including graph-based, tensor decomposition, and probabilistic graphical models have been extensively used [22]. Additionally, to cope with complex correlations between labels, researchers have employed strategies like the attention mechanisms and self-attention mechanisms. In this study, the self-attention mechanism is used to address the complex interactions between numerous labels.

By reading the literature and examining the emotional stances of information producers, this study uses a multi-label classification engine to categorize each text dataset and four categories are determined under the labels involved: pricing information, appearance information, quality information, and useless information. As mentioned previously, our study uses specific keywords to as the search term crawl textual information regarding ceramics. However, since ceramic text data may span multiple categories, the keywords alone cannot be a reliable method for annotation. Therefore, data annotation standards were prepared for expert annotators to follow and the standards are kept consistent throughout the task. Table 4 shows the annotation standards and keywords involved when the dataset is constructed.

Table 4. The annotation standards and keywords for the Data set

Label Type	Criteria	Keywords
Price Information	Information publisher's description of the price of ceramics	Inexpensive, cheap, affordable, bargain, discount, event, value, expensive, precious, valuable
Appearance Information	Descriptions of ceramic colors, patterns, shapes, styles, capacities, sizes, etc.	Exquisite, good-looking, creative, cute, right size, enough capacity, brightly colored, clear pattern, unique shape
Quality Information	Description of ceramic workmanship, materials, quality, etc.	Exquisite, fine workmanship, smooth, thick, delicate feel, practical, heat preservation, heat insulation, too thin, rough workmanship, defective, poor quality
Useless information	Does not deal with the above three classifications of ceramics and non-ceramic-related content	—

The data annotation standards are as follows [23]:

(1) Price information also deals with the reasonableness and consistency of the ceramic pricing. Price information can be categorized as the importance reviewers accord the value of ceramic products while making a purchasing decision.

(2) Appearance information also includes that about whether the ceramic is beautiful, colorful, or delightful, and so on. Appearance information can be categorized as different aesthetic tendencies and personalization needs of the commentator.

(3) Quality information is also including with whether the ceramics are produced with outstanding craftsmanship and reliable materials. Quality information can be categorized as the commentator's requirements for ceramic fabrication, durability, and structural stability.

2.3. Annotation guidelines

The following principles are observed in the annotation process:

(1) Three professional annotators in the field of ceramics are to be selected. The three annotators should all be experts with years of experience in the ceramics industry.

(2) One complete dataset will be assigned to at least two of the three annotators who would mark the ceramic texts with 1-3 emotion labels.

(3) The annotator's comments will be scrutinized and analyzed after every 500 tweets to ensure credibility and correctness.

(4) The annotators were required to interpret properly in accordance with different contexts so as for us to obtain a broader range of various annotation data. This will improve the generalization capability of well-trained models in real-world settings and environments.

(5) Even if only one category is labeled differently by the first two annotators, a serious contradiction may occur; and then the third annotator is to resolve the conflict over the contradictory datasets.

(6) If the contradiction cannot be resolved, it will be put to the vote; the label which wins receives more than two-thirds of the total valid votes will be approved as the one for the final data category.

In this study, the data are labeled with the values 0, 1, and -1, of which 1 denotes positive information about the label in the sample, -1 denotes negative information about the label in the sample, and 0 denotes null information about the label in the sample. It is important to note that the information labeled useless is mutually exclusive with the other three types of information and cannot appear as any one of the other three types, whereas the other three types of labels can used repeatedly for the same piece of information in a sample. This annotation method can help us to better understand and categorize ceramic text information and identify the emotions embedded in it.

2.4. BERT model

BERT is a contextualized word representation model that is based on a masked language model and pretrained using bidirectional transformers [24]. BERT obtains state-of-the-art performance in most NLP tasks, and it is an effective tool for tackling various NLP tasks [27]. Liu implemented extractive text summarization based on BERT to retain essential information in the text [28]. Huang et al. applied BERT to Chinese word segmentation to improve segmentation accuracy [29]. Goldberg tested BERT's syntactic analysis capabilities [30]. Masahiro Kaneko et al. categorized the syntactic error detection task as a sequence annotation task and achieved a significant performance improvement by using BERT [31].

The main structure of the BERT model consists of a 12-layer Transformer and a multi-head self-attention mechanism [34]. Each Transformer layer consists of multiple self-attentive sublayers and feedforward neural network sublayers. BERT, can successfully capture the correlation between words and contextual information. BERT, which is based on the bi-directional Transformer learning mechanism, can possesses great transfer learning capabilities without a sizable training corpus, saving time and improving work efficiency. In the pre-training process of BERT, the input consists of Word Embeddings, Segment Embeddings, and Position Embeddings. At the beginning of each sentence, a special marker [CLS] is used to mark the beginning of the sentence [35]. The last layer of hidden states corresponding to this position can be considered as a semantic representation of the whole sentence and it can be used for downstream tasks such as text categorization. Every two sentences input is separated using the [SEP] flag. This inputting mode fully utilizes the information of words, fragments, and locations, allowing BERT to better capture the correlations between words and the semantic representation of sentences. This design greatly enhances the performance of the BERT model in various text-processing tasks. The inputs of the BERT model are illustrated in Fig.2.

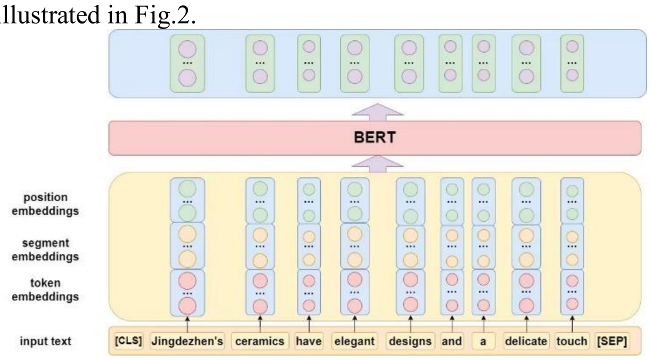

Figure 2. Inputs for the BERT-based model

The pre-training phase of the BERT model includes the MLM (Mask Language Model) and NSP (Next Sentence Prediction) tasks [37]. In the MLM task, the BERT model is used to predict the replacement token after a portion of the training corpus is marked with [MASK] strings [38]. This approach is equivalent to the practice of filling in the blanks when learning English, which can enhance the model's comprehension of the connotation beyond the between words [39]. The MLM tasks of the BERT model are illustrated in Fig. 3. In contrast, the BERT model's job in the NSP assignment is to take two sentences as input and determine whether or not there exists a local contextual link between them [40]. This task enables the BERT model to understand relationships between sentences, and then it can answer question automatically and perform linguistic reasoning naturally, etc. [41]. The combination of the MLM and NSP tasks further enhances the performance of the BERT model in various text-processing tasks [42].

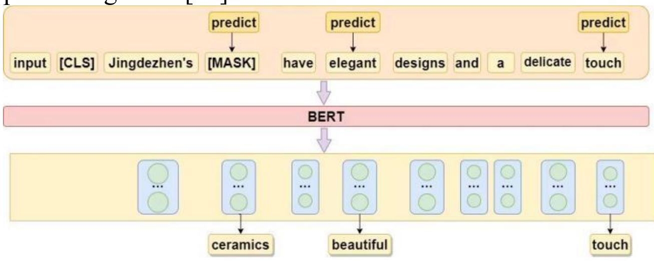

Figure 3. Inputs for the BERT-based model

2.5. Metrics for measuring the prediction quality of multi-label systems

Some commonly evaluation methods are Coverage error, Labeled ranking average percentage (LRAP), Ranking loss, Multi-label cross-entropy loss function $\mathrm{H}(y,\hat{y})$, etc. and can be formulated as:

$$Ranking\ loss = \frac{1}{n_s}\sum_{i=0}^{n_s-1}\frac{1}{\|y\|_0(n_l-\|y_i\|_o)}\left|\left\{(k,,l): \widehat{f_{ik}} \leq \widehat{f_{ij}}, y_{ik}=1, y_{il}=0\right\}\right| \quad (1)$$

$$\boldsymbol{H(y,\hat{y}) = -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{l} y_{ij}\, log(\hat{y}_{ij}) + (1 - y_{ij})\, log(1-\hat{y}_{ij})} \quad (2)$$

$$\boldsymbol{coverage\ error = \frac{1}{n_s}\sum_{i=0}^{n_s-1} \max_{j:y_{ij}=1} rank_{ij}} \quad (3)$$

$$\boldsymbol{LRAP = \frac{1}{n_s}\sum_{i=0}^{n_s-1}\frac{1}{\|y_i\|_0}\sum_{j:y_{ij}=1}\frac{|L_{ij}|}{rank_{ij}}} \quad (4)$$

Where n is the number of samples and l is the number of label types, ij, ik and il represent the jth, kth, and lth labels of the ith sample respectively. $\hat{f}$ represents the label prediction score, $rank$ represents label ranking, and $rank_{ij} = \left|\left\{k: \widehat{f_{ik}} \geq \widehat{f_{ij}}\right\}\right|$ represents the number of labels scoring higher than or equal to label ij. The statistic of outperforming a specific label is represented by L, where $L_{ij} = \left|\left\{k: y_{ij}=1, \widehat{f_{ik}} \geq \widehat{f_{ij}}\right\}\right|$ represents the number of labels ik in all samples scoring higher than label ij. y_{ij} is a binary indicator vector of the true label that shows whether sample i belongs to label j and $\hat{y}_{ij}$ is the chance that sample i will appear on label j according to the model. To determine the number of non-zero elements, we employ the L_o paradigm. Coverage error is used to assess prediction accuracy and it indicates the average number of true labels that need to be included. Ranking loss takes into account the average number of incorrect tag pairs weighted by the inverse of correct and incorrect tags, with values closer to 0 indicating better results.

3. Profile of Data and Results

3.1. Case data sets

The research object of this study is the text tagged with "Jingdezhen Ceramics" on the Microblog platform from 2009 to 2022. Since the ceramic exhibition has attracted extensive and enduring attention throughout the whole planning phase, it is easy to find text data as training samples. Through data collection, we discovered 168 postings on ceramics at the microblogging site and more than 500 people with usernames including terms associated with ceramics. A total of 10,154 samples are collected as the raw data with the procedure. 7564 cleaned samples are left after manual screening of data quality and deletion of the data unrelated to the experiment. In this instance, 40% of the data was used as the test data set and 60% of the data for the training set [43].

3.2. Sample labeling and statistics

In this study, a total of 7564 data samples were labeled, and the process included multiple rounds of labeling and review by multiple people. Each training sample includes preprocessed data, raw text data, and a list of labels, the last of which is used to record the categories of information that the text data falls in.

For instance, if the raw sample data is "Saturdays · Jingdezhen is lively and bustling, a lot of creative and high-quality ceramics are expensive, Jingdezhen · Saturday Pottery." To begin with, we preprocessed the text to get rid of items like punctuation and degree adverbs. Then, we label the Pricing-related information as "-1," appearance information as "1," and quality information as "1" through the use of the keywords "very expensive," "creative," and "good quality" included in the data. In addition, we have given the category of useless information the value "0" because the content of such comment content pertains to ceramics and is not worthless information.

As a result, the list of labels for this data is [-1,1,0,0], suggesting that the review's content is comprised of both good and negative information about prices and appearances. We obtained a carefully chosen and labeled high-quality dataset after completion of the labeling work, which offers a solid foundation for the follow-up study and analysis. A sample depiction of the training data is illustrated in Table 5.

Table 5. Some samples of training data

Raw text data	Preprocessed text data	P	A	Q	U
Each piece of Jingdezhen pottery is exquisite, and I was astounded by the store's skill and the high cost of each work of art~! 😂	Each piece; Jingdezhen; pottery; exquisite; astounded; store's; skill; high cost; work of art	-1	1	1	0
I bought a crooked handmade plate in Jingdezhen that was inexpensive and looked good!	Bought; crooked; handmade; plate; Jingdezhen; inexpensive; looked; good;	1	1	-1	0
Saturday - Jingdezhen has a lot of creative, good quality ceramics very expensive, Jingdezhen - Saturday Pottery	Saturday; Jingdezhen; a lot of; creative; good quality; ceramics; very; expensive; Jingdezhen; Saturday; Pottery	-1	1	1	0

*Notes: P represents price, A for appearance, Q for quality, and U for useless.

According to the distribution of sample numbers in the benchmark data sets shown in Fig 4, there is an imbalance in the number of different labels concerning the training samples. Specifically, the number of appearance and quality information is higher, about two to three times the number of price tags. The imbalance may be caused by the following reasons:

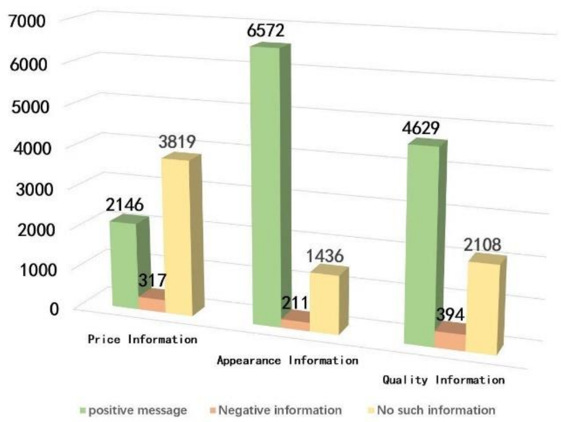

Figure 4. Inputs for the BERT-based model

When an individual comes into contact with a ceramic work, their initial impression is often strongly influenced by the appearance of the work. In addition, because of the differences in understanding and preferences of beauty, individual aesthetic standards will be different. As a result, consumers are more inclined to choose ceramics that meet their own aesthetic standards, which increased the number of labels for appearance-related information.

Secondly, the reason for a higher amount of quality information is that always attach more importance to the quality and workmanship of ceramic products. Good quality and fine workmanship ensure that ceramic products are strong, wear-resistant, and durable, which may meet the needs of reviewers and the expectations of users (or purchasers). In addition, the ceramics' quality and workmanship play a crucial role in maintaining and realizing their full value. High-quality ceramics are typically expensive and normally have as much market value as collectibles. As a result, more attention will be paid to the quality, which has led to an increase in the amount of information under the quality label. In-depth analysis of the imbalance between the amounts of information under the different labels can help us better understand people's concerns and preferences for ceramic products.

Table 6. Joint probability distributions for different label combinations

Tag-team	Number of labels	Joint probability	Tag-team	Number of labels	Joint probability
(0,1,1,0)	2600	0.344	(1,2,0,0)	3	0
(0,1,2,0)	115	0.015	(1,2,1,0)	2	0
(0,2,1,0)	19	0.003	(2,0,1,0)	12	0.002
(0,2,2,0)	27	0.004	(2,0,2,0)	6	0.001
(1,0,1,0)	271	0.036	(2,1,0,0)	46	0.006
(1,0,2,0)	15	0.002	(2,1,1,0)	33	0.004
(1,1,0,0)	323	0.043	(2,1,2,0)	5	0
(1,1,1,0)	859	0.114	(2,2,0,0)	1	0
(1,1,2,0)	21	0.003	(2,2,2,0)	3	0
(0,1,1,0)	2600	0.344	(1,2,0,0)	3	0

Approximately 58.4% (4,420 samples) of the total 7,564 samples can be marked by two or more labels. These multi-labeled samples can be grouped under 18 different conditions, which displays the complex relationships between the contents of various types of information. Our further investigation revealed a clear imbalance in the joint probability distribution of the various label combinations, particularly the excessively high combination probability of the positive appearance information and positive quality information. The joint probability distributions of the various label combinations are shown in Table 6. These findings demonstrate the conclusion that publishers have favorable opinions of the appearance and craftsmanship of ceramics, which is consistent with their expectations.

Besides, the probability of all tag combinations without positive appearance information appearing is relatively low. However, adding positive appearance information to the combination greatly increases the likelihood of the corresponding combination by more than 100 times. This means that positive appearance information appears frequently in published content, which may be related to the large number of appearance information tags in the sample. It is important to note that too many appearance information tags may affect the extraction effect of other tag content.

In conclusion, there are many instances of positive appearance information in published content, which may affect the extraction of other labeled content when there are many labels. These findings offer helpful references for learning more about the significance and classification of the information linked to ceramics.

3.3. Model parameters and experimental results

After the model has been trained, the hyperparameters must be adjusted to produce the best results [44]. Parameters such as learning rate and batch size are key parameters that need to be tuned [45]. By contrasting the model's accuracy across numerous rounds of experiments on each set of parameters, the optimal combination of parameters is determined. After analyzing the operation results of all the parameter combinations, the set of parameters with the highest accuracy is selected as the optimal solution. This set of parameters is considered to be the optimal parameter combination for the adjusted model. This strategy offers an effective way to improve model accuracy.

In the experiment, we adjusted the hyperparameters of the model by setting the learning rate to three values: 0.00001, 0.00001, and 0.00005, and setting the batch size of data processing to 16 and 32. After several rounds of experiments, we found that with an initial learning rate of 0.00005 and a batch size of 32, the model performs poorly with an accuracy of 0.64 and a loss value of 1.25. Further analysis suggests that a large batch size may result in the model unable to understand features of different samples. Therefore, with other parameters unchanged, we try to regulate the data processing batch size to 16, and then the experimental results show that the accuracy is 0.67 and the loss value is 1.15. From this we can reach that a smaller batch size can provide more accurate gradient estimation, increase the generalization ability of the model, and thus effectively avoid the overfitting problem. We conducted several rounds of experiments and found that the accuracy of the model is higher when the data processing batch size is 16 than at a data-processing batch size of 32. The comparison of experimental results under different batch sizes is shown in Table 7.

Table 7. The comparison of experimental results under different batch sizes with learning rate of 0.0005

Batch size	Accuracy	Loss value
32	0.64	1.25
16	0.67	1.15

In the following experiment, we take into account the learning rate as a significant element impacting the correctness of the test set after making adjustments to the data processing batch size. We changed the learning rate for the experiment from 0.00005 to 0.00002 while keeping the data processing batch size fixed at 32. According to the experimental findings, the accuracy was 0.83 and the loss was 0.55. Compared to the learning rate of 0.00005, this is an improvement of 0.20. Again, when adjusting the learning rate to 0.00001, the accuracy was 0.82, slightly lower than when the learning rate was 0.00002. Based on a data processing batch size of 32, we performed multiple tests with a learning rates of 0.00001, 0.00002, and 0.00005 respectively. According to the experimental findings, the model with a learning rate of 0.00002 has somewhat better accuracy than the model with a learning rate of 0.00001. This suggests that an ideal answer is not always produced at a lower learning rate. Instead, a higher learning rate helps the model to jump out from the local optimal solution and get into a more thorough exploration in the space of loss functions for the globally optimal solution. The comparison of experimental results under different learning rates is shown in Table 8.

Table 8. The comparison of experimental results under different learning rates with batch size of 32

Learning rate	Accuracy	Loss value
0.00005	0.64	1.25
0.00002	0.83	0.55
0.00001	0.82	0.56

The performance of the model can also be enhanced by properly sizing the processing data batch. Therefore, we changed the learning rate for the experiments to 0.00001, 0.00002, and 0.00005, respectively, and set the processed data batch size to 16. The experimental results show that with a learning rate of 0.00005, the model performance is significantly worse compared to other combinations of learning rates and is less accurate. The accuracy and loss values of the training set with different initial learning rates are shown in Fig. 5. In contrast, when the learning rate is 0.00002, the model predicted with the highest accuracy of more than 0.9 and with the lowest loss value of less than 0.4. In addition, there is little difference in the accuracy when the learning rates are set to 0.00001 and 0.00002, but the model performs better when the learning rate is 0.00001, allowing for a more fine-grained search of the parameter space. However, selecting a more appropriate learning rate can produce more effective experimental outcomes. We select the parameter combination with a batch size of 16 and a learning rate of 0.00002 as the model's optimal parameter combination after a careful consideration of the findings of numerous experiments.

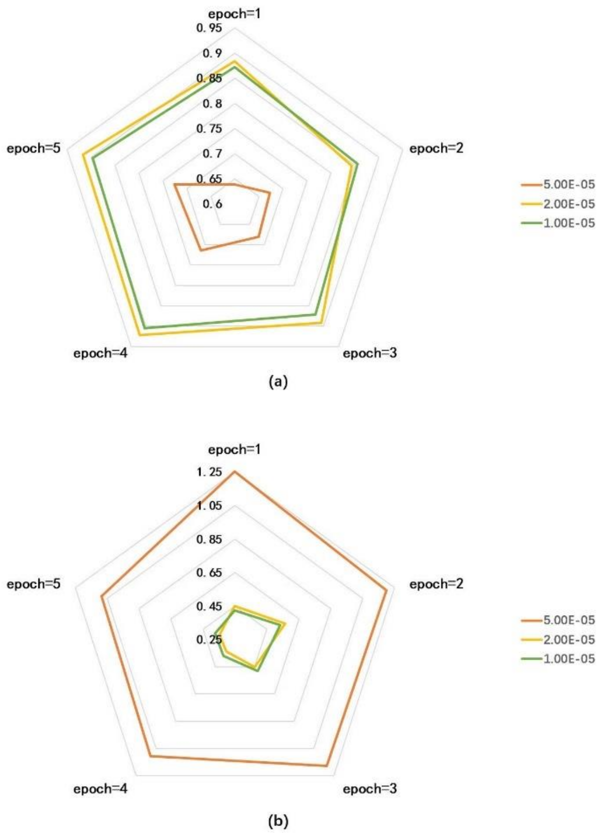

Figure 5. (a) Accuracy of the training set with different initial learning rates; (b) Training set loss values for different initial learning rates

3.4. Some notes

The BERT model has been pre-trained on a larger dataset to learn the more general linguistic features and has been fine-tuned to further adapt to the specific ceramic microblog text classification task. The model can learn all facets of contextual information in a sentence while simultaneously learning more in-depth semantic information thanks to its internal Transformer mechanism. Nevertheless, the BERT model has to be updated and enhanced in several areas.

1. Training blind spots, which is a common problem in machine learning and deep learning networks. The BERT model is unable to carry out accurate identification when the categories of some samples are not well distinguished in the vector space, and the presence of such samples causes the model's capacity for generalization to decline.

2. Catastrophic forgetting, which is a frequent phenomenon in deep learning. Deep learning networks may forget previously learned information as they learn new information, which is what distinguishes deep learning from advanced biology. When the BERT model goes through adaptive training, it may forget some of the information it learned during pre-training; in certain situations, this may cause the model to forget its prior information.

3. Overfitting problems, which means that the model overfits the dataset. The overfitting issue is caused by a variety of factors, including noisy samples in the dataset, an overly complex model with insufficient samples at once, and even over-training of the model. Due to the BERT model's extreme power and intricate internal structure, overfitting issues can readily arise while fine-tuning BERT.

Each of the aforementioned issues has a direct impact on how well the model performs in text categorization tasks, which lowers classification accuracy. Including some optimization tactics into the BERT model during training to boost the classification accuracy is an effective way to solve the above problems.

4. Conclusion

Social media platforms are continually changing, and their significance for the release and diffusion of information has made them crucial for analyzing market trends and public opinion in real time. How to effectively use social media information for text categorization has become one of the focuses of research. Currently, text categorization mainly uses binary and triple classification methods. However, multi-label classification is more suitable for the task of labelling social media messages that contain rich and diverse content and involve topics that may belong to multiple categories of information.

In this study, a dataset for sentiment analysis of ceramics is created for the first time. This dataset consists of the text information tagged with "Jingdezhen Ceramics" on the Microblog platform from 2009 to 2022. The study's dataset can be downloaded for free from GitHub (https:// github.com/fucrff/ a-dataset- for-sentiment- analysis- of-ceramics). This study has proposed a novel method for text recognition and classification of microblogging information related to Jingdezhen ceramics by fusing the Bert model and multi-label classifiers. Information is categorized according to the perception of ceramics held by the review publisher, models are constructed and trained with fine-tuning on a training set, and validated on a test set. We made the following deductions from the experiments' analysis:

1. The current study explores the application of the BERT model to the analysis of ceramic text information. The pre-training task makes use of the self-attention mechanism from the Transformer model's encoder component. The text categorization job is trained by fine-tuning after the pre-training phase is finished, which mostly entails adjusting the hyper-parameters to increase classification accuracy.

2. The BERT model was successfully used for the analysis of ceramic text data. By gathering pertinent ceramic text data from microblogs, preprocessing it, and manually labelling it, the quality of the dataset and the accuracy of the labels are guaranteed. The BERT pre-trained model was used for migration learning, and we verified it on the test set after fine-tuning the training on the training set. The model performs exceptionally well in terms of accurate classification and information extraction from ceramic data. It can also make judgments about the categories of ceramic text data based on its acquired knowledge.

3. Analyzing ceramic text data contributes to text mining efforts. Text comprehension and feature extraction accuracy in ceramic text data analysis are enhanced by the use of the BERT model. Simultaneously, it broadens the scope of text mining research and the methodologies' applications while enhancing the precision and comprehensiveness of ceramic text data analysis. Additionally, the model offers more in-depth perceptions of the textual data of the ceramics sector, offering powerful support for decision-making and product enhancement in the sector.

The study's findings indicate that the multi-label classification model has a wide range of potential applications in the classification of social media content. In addition, our work offers fresh approaches and strategies for handling and examining data from social media.

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), Scientific and Technological Research Projects of Jiangxi Provincial Department of Education (No. GJJ2201034) and Research Projects of the Provincial Department of Education in 2022 (GJJ2201039).

References

[1] Kaliyar, R.K., A. Goswami, and P. Narang, FakeBERT: Fake news detection in social media with a BERT-based deep learning approach. Multimedia tools and applications, 2021. 80(8): p. 11765-11788.

[2] Yuan, L., et al., Transfer learning for hate speech detection in social media. arXiv preprint arXiv: 1906.03829, 2019.

[3] Gundapu, S. and R. Mamidi, Transformer based automatic COVID-19 fake news detection system. arXiv preprint arXiv:2101.00180, 2021.

[4] Li, Y., Application of Text Classification Technology in the Ceramic Industry. Modern Computer (Professional Edition), 2010(14): p. 60-63.

[5] Yanan, Z., 'Moving with Emotion'—A Study on the Application of Emotional Elements in Daily Ceramic Design. 2013.

[6] Yun, Y., et al., A Decorative Pattern Classification Model for Ancient Ceramics Based on Improved E-EfficientNet. Laboratory Research and Exploration, 2023. 42(02): p. 129-134+186.

[7] Xinping, Y., Application of Data Mining in the Chronology of Ancient Ceramics and Classification of Ceramic Raw Materials. 2023.

[8] Song, R., et al., Label prompt for multi-label text classification. Applied Intelligence, 2023. 53(8): p. 8761-8775.

[9] Xiaofeng, L., L. Huili, and M. Yinglong, HE-HMTC: A hybrid embedding-based text representation for Hierarchical multi-label text classification. Software Impacts, 2022. 14.

[10] Yuhao, Z., Research on Short Text Classification of News Based on BERT. 2021.

[11] Chen, T., et al., The lottery ticket hypothesis for pre-trained bert networks. Advances in neural information processing systems, 2020. 33: p. 15834-15846.

[12] Wei, L., et al., Prompt Tuning for Multi-Label Text Classification: How to Link Exercises to Knowledge Concepts? Applied Sciences, 2022. 12(20): p. 10363.

[13] Ming, S. and L. Yanlong, Application and Optimization of BERT in Sentiment Classification of Weibo Short Texts. Journal of Miniaturization for Technology 2021. 42(04): p. 714-718.

[14] Luo, L., et al., An attention-based BiLSTM-CRF approach to document-level chemical named entity recognition. Bioinformatics, 2018. 34(8): p. 1381-1388.

[15] Zhang, T., et al., Revisiting few-sample BERT fine-tuning. arXiv preprint arXiv:2006.05987, 2020.

[16] Chen, Z. and T. Qian. Transfer capsule network for aspect level sentiment classification. in Proceedings of the 57th annual meeting of the association for computational linguistics. 2019.

[17] González-Carvajal, S. and E.C. Garrido-Merchán, Comparing BERT against traditional machine learning text classification. arXiv preprint arXiv:2005.13012, 2020.

[18] Jeong, C., et al., A context-aware citation recommendation model with BERT and graph convolutional networks. Scientometrics, 2020. 124: p. 1907-1922.

[19] Zeng, M., et al., Musicbert: Symbolic music understanding with large-scale pre-training. arXiv preprint arXiv:2106.05630, 2021.

[20] Yanni, M., Research on Multi-label Sentiment Classification Problem Based on Machine Learning. Computer and Information Technology, 2023. 31(05): p. 12-15.

[21] Wenbiao, Z., Research on Multi-label Classification Algorithm Based on Label Correlation Analysis. 2023.

[22] Xuegang, L. and L. Junrui, A Water Quality Prediction Model Method Based on Tensor Feature-GRU and Multi-head Self-attention Mechanism. Journal of Panzhihua University, 2023. 40(05): p. 89-96.

[23] Ashraf, N., et al., Multi-label emotion classification of Urdu tweets. PeerJ Computer Science, 2022. 8: p. e896.

[24] Jian, D., L. Junhua, and C. Jingang, An Improved HiNT Text Retrieval Model Using BERT and Coverage Mechanism. Journal of Intelligent Systems: p. 1-8.

[25] Huang, Y.-S. and Y.-H. Yang. Pop music transformer: Beat-based modeling and generation of expressive pop piano compositions. in Proceedings of the 28th ACM international conference on multimedia. 2020.

[26] Hsiao, W.-Y., et al. Compound word transformer: Learning to compose full-song music over dynamic directed hypergraphs. in Proceedings of the AAAI Conference on Artificial Intelligence. 2021.

[27] Lu, Z., P. Du, and J.-Y. Nie. VGCN-BERT: augmenting BERT with graph embedding for text classification. in Advances in Information Retrieval: 42nd European Conference on IR Research, ECIR 2020, Lisbon, Portugal, April 14–17, 2020, Proceedings, Part I 42. 2020. Springer.

[28] Liu, Y., Fine-tune BERT for extractive summarization. arXiv preprint arXiv:1903.10318, 2019.

[29] Huang, W., et al., Toward fast and accurate neural chinese word segmentation with multi-criteria learning. arXiv preprint arXiv:1903.04190, 2019.

[30] Goldberg, Y., Assessing BERT's syntactic abilities. arXiv preprint arXiv:1901.05287, 2019.

[31] Kaneko, M. and M. Komachi, Multi-head multi-layer attention to deep language representations for grammatical error detection. Computación y Sistemas, 2019. 23(3): p. 883-891.

[32] Delobelle, P., T. Winters, and B. Berendt, Robbert: a dutch roberta-based language model. arXiv preprint arXiv:2001.06286, 2020.

[33] Devlin, J., et al., Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018.

[34] Prasanna, S., A. Rogers, and A. Rumshisky, When BERT plays the lottery, all tickets are winning. arXiv preprint arXiv:2005.00561, 2020.

[35] Chuag, L., BERT's Application in Text Analysis. 2020.

[36] Xie, S., et al., PALI at SemEval-2021 task 2: fine-tune XLM-RoBERTa for word in context disambiguation. arXiv preprint arXiv:2104.10375, 2021.

[37] Bhasuran, B., BioBERT and Similar Approaches for Relation Extraction, in Biomedical Text Mining. 2022, Springer. p. 221-235.

[38] Clark, K., et al., Electra: Pre-training text encoders as discriminators rather than generators. arXiv preprint arXiv:2003.10555, 2020.

[39] Lianxi, W., et al., Research on Recognition and Feature Distribution of ASEAN-related Public Opinion towards China: Analyzing Mainstream English and Chinese Media. Journal of Intelligence, 2022. 41(08): p. 94-101.

[40] Joshi, M., et al., Spanbert: Improving pre-training by representing and predicting spans. Transactions of the association for computational linguistics, 2020. 8: p. 64-77.

[41] Yang, Z., et al., Xlnet: Generalized autoregressive pretraining for language understanding. Advances in neural information processing systems, 2019. 32.

[42] Xu, W., et al. Symmetric regularization based bert for pair-wise semantic reasoning. in Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. 2020.

[43] Chuan, C.-H. and D. Herremans. Modeling temporal tonal relations in polyphonic music through deep networks with a novel image-based representation. in Proceedings of the AAAI Conference on Artificial Intelligence. 2018.

[44] De Bruyn, M., et al. Bart for knowledge grounded conversations. in Proceedings of KDD Workshop on Conversational Systems Towards Mainstream Adoption (KDD Converse'20). ACM, New York, NY, USA. 2020.

[45] McCarley, J., R. Chakravarti, and A. Sil, Structured pruning of a bert-based question answering model. arXiv preprint arXiv:1910.06360, 2019.