PLOS ONE

RESEARCH ARTICLE

A study of text classification algorithms for live-streaming e-commerce comments based on improved BERT model

Rong Zhou[1], **Qing Shen**[2*], **Huafeng Kong**[2]

1 Faculty of Business and Economics, University of Malaya, Kuala Lumpur, Malaysia, **2** Department of Information Engineering, Wuhan Business University, Wuhan, China

* shenqing@wbu.edu.cn

Check for updates

OPEN ACCESS

Citation: Zhou R, Shen Q, Kong H (2025) A study of text classification algorithms for live-streaming e-commerce comments based on improved BERT model. PLoS ONE 20(4): e0316550. https://doi.org/10.1371/journal.pone.0316550

Editor: Weiqiang (Albert) Jin, Xi'an Jiaotong University, CHINA

Received: July 31, 2024

Accepted: December 12, 2024

Published: April 22, 2025

Copyright: © 2025 Zhou et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: This work was supported by Ministry of Education Industry-university Cooperative Education Project (230804144213810), Hubei Provincial Department of Economy and Information Technology Scientific Research Project (GXCZ-C-23130484),

Abstract

As e-commerce live streaming becomes increasingly popular, the textual analysis of bullet comments is becoming more and more important. Bullet comments is characterized by its brevity, diverse content, and vast quantity. Faced with these challenges, this study proposes an improved BERT model based on a hierarchical structure for classifying e-commerce bullet comments. First, a parent class BERT model is trained to categorize bullet comments into six designated categories (parent categories). Subsequently, subclass BERT models are trained to classify bullet comments into subcategories. The model combines BERT's profound semantic comprehension with the closely categorized capabilities of the hierarchical structure. Empirical evidence shows that the proposed model significantly improves classification accuracy and efficiency, aiding in further analysis of bullet comments, extracting valuable information, and achieving effective marketing.

1. Introduction

The digital revolution has catalyzed a surge in the popularity of e-commerce live streaming, which has quickly become a focal point of global attention and a new frontier in online shopping [1]. This innovative approach to retail has been particularly notable in China, where the live streaming e-commerce industry has experienced explosive growth [2]. By 2022, the live streaming e-commerce industry in China had reached a new peak, with a transaction volume of 3,487.9 billion yuan [3]. Live streaming e-commerce enables viewers to send and display real-time comments on the live video stream [4]. These comments, known as 'bullet comments' or 'danmu', appear on the video in a scrolling, floating, or stationary manner, offering a novel way for viewers to engage with the live broadcast [5]. Viewers can post comments and feedback in real-time during the live stream, allowing hosts to quickly respond to audience needs and questions by monitoring these bullet comments [6]. Furthermore, other viewers can glean product information and purchasing experiences from these comments, which can inform their own buying decisions [7]. The application of bullet comments in live streaming e-commerce not only enriches the presentation of live broadcasts but also introduces new interactive and marketing opportunities for the e-commerce industry. Consequently, it is

Hubei Province Education Science Planning Project (2023GB163) and Outstanding Young and Middle-aged Scientific and Technological Innovation Teams in Higher Education Institutions of Hubei Province (Number:T2023038). The funders had a role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

essential to perform text classification on these comments to gain a deeper understanding of viewer preferences, thereby enhancing marketing efficiency.

Within the realm of text classification techniques, the prevailing approaches encompass classification utilizing semantic dictionaries, conventional machine learning methodologies, and deep learning strategies. Meng, Duan [1] employed a sentiment dictionary to analyze viewer comments in live-streaming e-commerce, a method that, while simple to implement, has inherent limitations in accuracy. Dragoni, Federici [8] explored the application of traditional machine learning for comment categorization; however, this technique faces challenges when applied to extensive datasets. The BERT model from deep learning has garnered widespread recognition for its superior performance in addressing such challenges [9, 10].

Currently, a significant portion of scholarly research is conducted using flat text classification techniques, as evidenced by the works of Cao, Sun [11] and Su, Cheng [12]. These methods involve directly categorizing text data into a predefined set of non-hierarchical categories, without considering the potential for nested or hierarchical relationships among these categories [11, 12]. Despite various enhancements made, the accuracy of these flat classifications has not seen substantial improvement. In contrast, Ma, Liu [13] introduced a hierarchical classification approach and, through a series of five experiments, demonstrated that this method surpasses flat classification in precision. Hierarchical classification organizes categories into a structured hierarchy, where each category functions as a node within a tree-like architecture [14]. This structure allows for categories to have subcategories (child nodes) and a superior category (parent node), creating a directed acyclic graph (DAG) that facilitates a more nuanced and systematic approach to classification [15, 16]. Building upon these insights, this study proposes a hierarchical classification method to augment the BERT model. By incorporating a hierarchical structure into the classification process, we anticipate a more accurate text classification that can better capture the complexities and nuances of viewer feedbacks within the live-streaming e-commerce ecosystem.

Therefore, this study introduces an enhanced BERT model structured hierarchically for classifying e-commerce live broadcast bullet comments. Initially, a parent class BERT model is trained to categorize bullet comments into six predefined categories (parent categories). Following this, the subclass BERT models are trained to further classify the bullet comments into subcategories. This model leverages BERT's deep semantic understanding, complemented by the hierarchical structure's ability to tightly classify closely related categories. Empirical evidence demonstrates that the proposed model notably enhances classification accuracy and efficiency.

The contribution of this study is the construction of an innovative BERT model, based on a hierarchical structure. Through this new BERT model, researchers can extract valuable information from bullet comments with the characteristics of brevity, diverse content and vast quantity, which is important for understanding and interpreting viewers' preferences.

The remaining parts of this paper are organized as follows: Section 2 Related Work; Section 3 Method; Section 4 Experiments; Section 5 Results and Discussions and Section 6 Conclusions and Future work.

2. Related work

2.1 Text classification

The text classification process, a pivotal domain within Natural Language Processing (NLP), is utilized for a variety of applications such as Sentiment Analysis, Topic Labeling, Question Answering, Dialog Act Classification, and Natural Language Inference [17]. This methodology involves several key steps: obtaining original data, data preprocessing, feature extraction,

classifier application, and the generation of category outputs [18]. Original data is often sourced from platforms like Facebook, Twitter, and e-commerce websites, using Python for extraction [19]. Data preprocessing is a crucial step that includes data cleaning to remove noise, standardizing text categories (e.g., Chinese or English), employing word segmentation techniques, and filtering out stop words [20]. Feature extraction can be approached in various ways; traditional methods like the N-gram and TF-IDF are common, while deep learning often relies on automatic feature extraction. Classifiers such as SVM and SoftMax are employed to determine the text's final characteristics [8]. Text classification techniques encompass semantic dictionary-based methods, traditional machine learning, and deep learning strategies.

Semantic dictionary-based classification uses a specialized dictionary for text identification and categorization. This method involves text input, preprocessing, segmentation, training with the semantic dictionary, and classification based on established rules [9]. For example, Meng, Duan [1] used an emotional dictionary to analyze viewer comments on live streams, exploring emotional contagion. However, this approach, while simple, has limitations in accuracy [9].

Machine learning approaches involve training models on text classification tasks, extracting features from extensive text corpora, and predicting classifications. These methods are categorized into supervised, semi-supervised, and unsupervised learning [8]. Dragoni, Federici [8] employed an unsupervised method for classifying real-time comments. Despite their higher accuracy, machine learning methods may not be ideal for very large datasets [9].

Deep learning methods encompass a range of techniques from single neural networks to hybrid models, attention mechanisms, and pre-trained models [18]. Pre-trained models, like Google's BERT introduced in 2018, are a focal point of current research. They offer the advantage of capturing intricate lexical relationships and can be fine-tuned for excellent performance in specific tasks [21]. BERT uses a bidirectional mechanism for contextual word understanding and combines Word Piece embeddings with positional encoding [22]. Araci [23] achieved notable success with BERT for text classification due to its high accuracy and suitability for large datasets [9]. In the past few years, the BERT model has gained popularity for its effectiveness in classifying review text. For example, Su, Cheng [12] used an improved BERT model to classify social media comments, which helped in tracking the shifts in public sentiment. Similarly, Cao, Sun [11] applied an enhanced BERT model for sentiment analysis on reviews of agricultural products.

2.2 Hierarchical structure

The flat text classification method treats text data in a straightforward manner, ignoring any hierarchical structures [14]. Despite efforts to improve this method, there has not been a significant leap in classification accuracy. Several factors contribute to this challenge. Text classification models are highly complex with numerous parameters, indicating their proficiency in capturing linguistic nuances, which means that further modifications may yield only marginal performance improvements [15]. Additionally, the quality and size of the training data can limit the model's ability to generalize and enhance accuracy if the dataset lacks diversity or is not large enough [20]. Lastly, the complexity inherent in certain text classification tasks can be problematic [22]. This complexity may stem from the subjective nature of the text, ambiguities within it, or the need for domain-specific knowledge that may not be captured by models pre-trained on general text [24].

Ma, Liu [13] highlighted that traditional flat classification methods struggle with the vast volume and subtle distinctions between categories in real-world text classification scenarios.

To address this, they proposed a hierarchical classification method that is more adept at managing complex text classification challenges. Their experiments across five real-world datasets showed that hierarchical classifiers generally outperformed flat classifiers. Hierarchical classification methods offer a structured approach to text classification, particularly useful for tasks with large volumes of text and closely related categories [25]. These methods organize classes into a tree-like hierarchy, with nodes representing categories and edges showing the relationships between parent and child nodes, allowing for a more nuanced classification process [26].

Unlike flat classification systems, which assign texts to one or more categories at the same level and can lead to confusion with overlapping categories, hierarchical approaches provide clarity [27]. They establish a nested structure that organizes similar categories, enhancing classification precision [28]. As the number of texts and categories grows, flat models may become inefficient due to the curse of dimensionality [29]. Hierarchical classification, however, breaks down the classification problem into smaller, more manageable tasks, making it more effective for handling a larger number of categories [30]. It also allows for the addition of new categories without requiring a complete retraining of the model, offering flexibility for dynamic classification needs [31]. Furthermore, hierarchical classification improves the interpretability of the classification process by providing a clear sequence of decisions that led to a particular classification [32]. It also excels at distinguishing between closely related or overlapping categories by training the model to differentiate based on hierarchical relationships [33].

Considering the vast amount of bullet comments in live streaming e-commerce and the closeness of classification categories, to improve classification accuracy, this study employs a hierarchical classification approach.

3. Method

A BERT model based on hierarchical structure is proposed, referred to as HS-BERT. This is a tiered text classification method where categories are not flat but form a tree-like structure, with some categories being subcategories of others. The parent category is classified into N subcategories using the parent class BERT model, and each subcategory is further classified using a respective subclass BERT model. The framework of our HS-BERT method is illustrated in the Fig 1.

The first is Input Layer. The input layer receives the user's review sentence s, which is composed of a series of word elements expressed as

$$s = \left\{ w(s)_1, w(s)_2, \ldots, w(s)_n \right\} \tag{1}$$

Hierarchical layer is followed. There are two parts. 1) Top-level Classification: First, a BERT model is used to classify the text to determine which top-level parent category it belongs to. This step involves training a parent class BERT classifier that can identify the most likely parent category for the text. 2) Subcategory Classification: Once it is determined which parent category the text belongs to, the next step is to use subclass BERT models to further classify the text into the respective subcategories.

The BERT model is divided into six sections. The first is Word Embeddings. Initially, each word in the sentence is converted into a word embedding vector.

$$E = Embedding(words) \tag{2}$$

words refers to the sequence of words in a sentence. E refers to the word embedding matrix, where each word is mapped to a fixed-dimensional vector space.

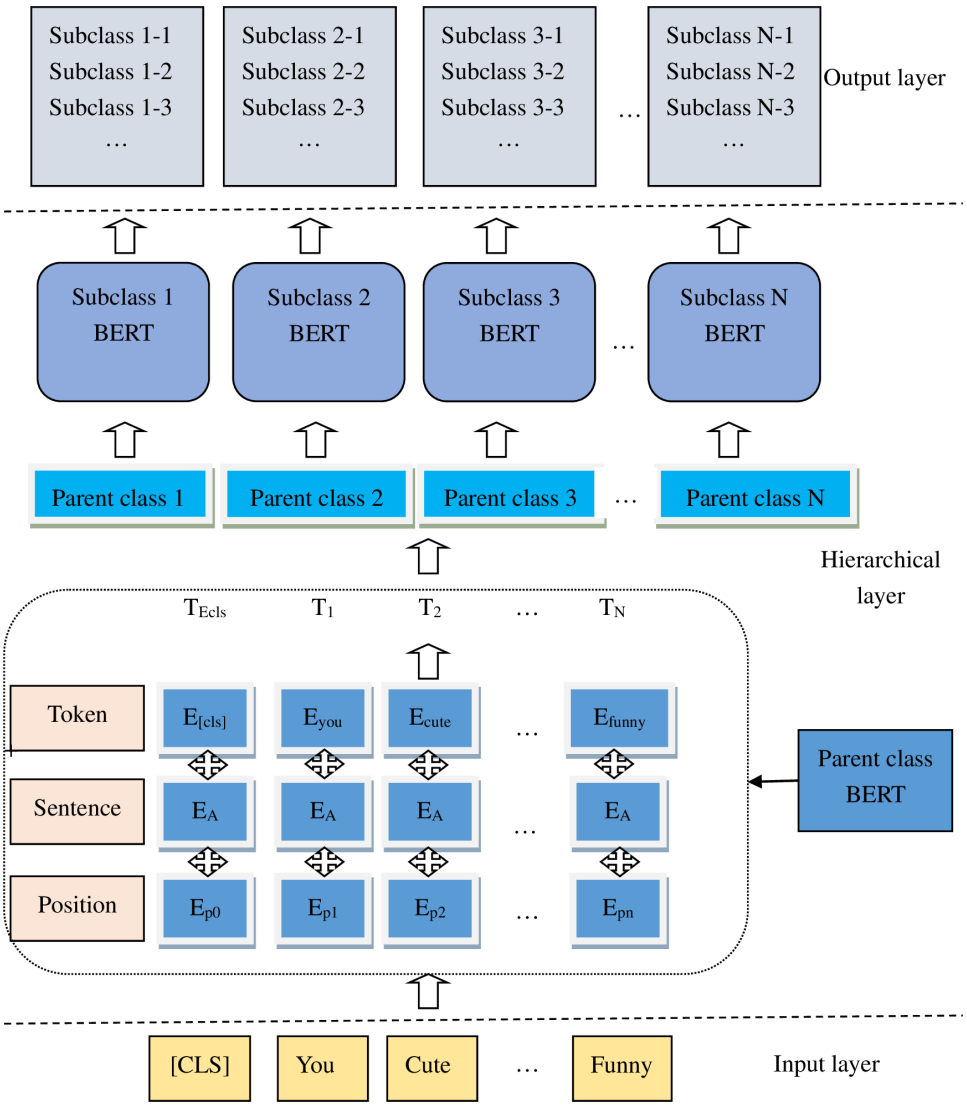


Fig 1. The Structure of the Hierarchical BERT Model.

https://doi.org/10.1371/journal.pone.0316550.g001

The second is Positional Encoding. Positional encodings are added to the word embeddings to provide information about the position of the words in the sentence.

$$X = E + P \tag{3}$$

P represents the positional encoding, a vector of the same dimension as the word embedding, whose values are generated by sine and cosine functions. X is the input vector that combines word embeddings and positional information. The specific calculation method for positional encoding:

$$PE(pos,2i) = sin\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right) \tag{4}$$

$$PE(pos,2i+1)=cos\left(\frac{pos}{10000\frac{2i+1}{d_{model}}}\right) \tag{5}$$

pos refers to the position of the word in the sentence. i refers to the index of the dimension. d_{model} refers to the word embedding dimension of the model.

The third is Segment Embeddings. When processing segments, different embeddings are used to distinguish between the two sentences.

$$X_{seg}=Eembedding(segments) \tag{6}$$

$segments$ indicate the type of each sentence (e.g., the first sentence or the second sentence). X_{seg} is the type embedding for the segment.

The fourth is Muti-head Self-Attention. In each encoder layer, the self-attention mechanism allows the model to focus on multiple positions simultaneously when processing a sentence.

$$Q,K,V=XW^{Q},XW^{K},XW^{V} \tag{7}$$

$$Attention(Q,K,V)=Softmax\left(\frac{QK^{T}}{\sqrt{d_k}}\right)V \tag{8}$$

Q,K,V represent the query, key, and value matrices. W^{Q},W^{K},W^{V} are learnable weight matrixes. d_k is the dimension of the key vectors.

The fifth is Residual Connection and Layer Normalization. The output of each sub-layer (self-attention or feed-forward network) goes through a residual connection and layer normalization.

$$L=LN(A+X) \tag{9}$$

L is the output after layer normalization and residual connection. LN represents the layer normalization operation.

The last is Feed-Forward Neural Network. Each encoder layer also includes a feed-forward network for further processing the output of the self-attention layer.

$$F=\max(0,LW_1+b_1)W_2+b_2 \tag{10}$$

W_1,W_2,b_1,b_2 represent the learnable weights and biases of the feed-forward network.

The final layer is the output layer.

4. Experiments

4.1 Data acquisition and preprocessing

In this research, viewer comments on TikTok's Chinese version have been gathered. The decision to focus on China as the research subject is based on the swift growth of live-streaming e-commerce in the nation since 2016, along with a persistent rise in market penetration [1]. The platform of TikTok (Chinese version), which boasts 700 million daily active users in

China, is selected for the study due to its extensive reach among the country's populace of 1.4 billion [34].

Two data sets are selected in Table 1. Tea and Snacks datasets focus on bullet comments from tea and snacks sales from January 1, 2023, to April 18, 2023 and is publicly accessed on the Chanmama Data Platform (https://www.chanmama.com/). Data collection follows the Chanmama Data Platform's Service Agreement (https://www.chanmama.com/other/serviceAgreement.html) and Privacy Policy (https://www.chanmama.com/other/privacyAgreement.html).

This paper annotates viewer comments. The parent category annotation divides the reviews into 6 categories based on the work of Shen, han Wen [3], namely evaluation, inquiry, promotion, price, logistics, and influencer, followed by manual annotation. The subcategory annotation further classifies the parent categories based on existing literature. Specifically, the evaluation category is broken down into three subcategories: quality, packaging, and after-sales service [1]. The inquiry section is divided into six parts: variety, quantity, quality, origin, packaging, and price [35]. The promotion category is subdivided into three subcategories: scarcity promotions, monetary promotions, and non-monetary promotions [36]. The price category is split into two subcategories: expensive and cheap [1]. The logistics category is detailed into six subcategories: reliability, economy, empathy, timeliness, flexibility, and informativity [37]. Lastly, the influencer category is divided into four subcategories: professionalism [38], homogeneity [2], attraction [39], and interactivity [40]. Manual annotation is followed. The corresponding labels can be found in Table 2 and 3. The data is randomly sorted with 80% as the training set and 20% as the validation set.

4.2 Experimental environment

The experimental setup utilized computing hardware in the form of an NVIDIA GeForce RTX 3090 with 64GB of memory, and the deep learning framework employed was PyTorch [41]. The detailed specifications of the experimental environment are depicted in Table 4.

4.3 Model parameter setting

The relevant parameters of the model are shown in Table 5 based on the work of Shen, han Wen [3]. The selection of an optimizer is pivotal as it can markedly influence the outcomes of model training [42]. Guan [43] examined the performance of Adam and AdamW optimizers.

Table 1. Two datasets.

Dataset Name	Sample
Tea	16,224
Snack	12,979

https://doi.org/10.1371/journal.pone.0316550.t001

Table 2. Parent class classification label.

Parent class	Description
Evaluation	The text contains evaluations of product performance.
Inquiry	The text contains inquiries about product information.
Promotion	The text includes reactions to promotional strategies.
Price	The text includes feedback on product price.
Logistics	The text includes feedback on shipping services.
Influencer	The text includes reactions to live streaming influencers.

https://doi.org/10.1371/journal.pone.0316550.t002

Table 3. Subclass classification label.

Parent class	Subclass	Description
Evaluation	Quality	The text contains evaluations of the quality of the product, including durability, performance, materials and so on.
	Packaging	The text contains evaluations of the product packaging, concerning the aesthetics of the packaging, protection, etc
	After-sales	The text contains comments on after-sales service, such as return and exchange policies, customer service response speed and service quality
Inquiry	Variety	The text contains queries about product varieties
	Quantity	The text contains queries about the number of products
	Quality	The text contains inquiries about the quality of the product
	Origin	The text contains queries about the origin of the product
	Packaging	The text contains queries about product packaging
	Price	The text contains inquiries about the price of the product
Promotion	Scarcity promotions	The text contains responses to scarcity promotions such as limited-quantity, limited-time, etc
	Monetary promotions	The text contains responses to monetary incentives such as price concessions, discounts, and buy-one-get-one-free offers.
	Non-monetary promotions	The text contains responses to non-monetary incentives such as gifts, points, coupons, etc
Price	Expensive	The text contains feedback on high price
	Cheap	The text contains feedback on low price
Logistics	Reliability	The text contains evaluations of the reliability and safety of logistics services
	Economy	The text contains evaluations of logistics costs
	Empathy	The text contains the evaluation of customer service in the process of logistics service
	Timeliness	The text contains comments on the speed of logistics, including delivery and distribution
	Flexibility	The text includes evaluations of the flexibility of logistics services, such as the ability to specify delivery time or locations
	Informativity	The text contains evaluations of the transparency and frequency of updates regarding logistics information
Influencer	Professionalism	The text contains feedback on the influencer's expertise.
	Homogeneity	The text contains feedback on the influencer's homophily
	Attraction	The text contains feedback on the influencer's attractiveness
	Interactivity	The text contains feedback on the influencer's interactivity

https://doi.org/10.1371/journal.pone.0316550.t003

The findings indicated that AdamW, when integrated with weight decay (denoted as »), demonstrated superior efficacy in regularizing the model and curbing overfitting tendencies. Analyzing the impact of learning rates on the loss curve is essential for finding the optimal rate, ensuring model stability, and speeding up convergence. Monitoring epochs' effect on accuracy helps in tracking learning progress, detecting overfitting, and enhancing the model's generalization. Adjusting dropout rates may necessitate more epochs for learning, while adding transformer layers should be paired with an increase in hidden neurons to maintain representational power. Varying convolution sizes and kernel sizes allows the model to capture features at different scales.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{11}$$

Table 4. Experimental environment.

Projects	Configuration
Operating Platforms	CUDA 10.1
Operating System	Windows 11
Memory	64GB
Python Versions	Python 3.8.1
Pytorch Versions	Pytorch 2.2

https://doi.org/10.1371/journal.pone.0316550.t004

Table 5. Model parameter settings.

Hyper-parameters	Value
Learning rate	1e-5
Loss Function	Cross-Entropy Loss
Optimizer	AdamW
Batch size	32
Dropout	0.1
Convolution size	3×3
Kernel sizes	[3–5]
Epochs	10
Transformer layers	12
Transformer number of hidden neurons	768

To assess the effectiveness of the models, performance metrics including accuracy, precision, recall, and the F1 score are applied. In the context of classification, instances where positive instances are correctly identified are termed True Positives (TP), while those incorrectly classified as negative are known as False Negatives (FN). Conversely, cases where negative instances are incorrectly labeled as positive are referred to as False Positives (FP), and when negative instances are correctly identified, they are called True Negatives (TN).

https://doi.org/10.1371/journal.pone.0316550.t005

$$Precision = \frac{TP}{TP + FP} \tag{12}$$

$$Recall = \frac{TP}{TP + FN} \tag{13}$$

$$F1 = \frac{2 * Precison * Recall}{Precision + Recall} \tag{14}$$

5. Results and discussions

Since HE-BERT is structured hierarchically, following the work of Ma, Liu [13], we only need to obtain the accuracy, precision, recall, and F1 score of the subclass BERT models, as shown in Table 6. To facilitate comparison with other models, we take the average of these four metrics of the subclass BERT models.

The efficacy of the HS-BERT model is compared against baseline research. The model is evaluated in comparison with SVM, CNN, and BAYES models. Evaluation metrics include accuracy, precision, recall and F1 score. Comparison results are in Fig 2.

On the Tea dataset, the accuracy for SVM, CNN, BAYES, and HS-BERT are 91.67%, 90.16%, 86.07%, and 98.22%, respectively. Notably, HS-BERT outperformed the other models in terms of accuracy. In terms of precision, the respective rates are 91.42% for SVM, 90.38% for CNN, 87.68% for BAYES, and 95.41% for HS-BERT, with HS-BERT again showing

Table 6. Four Indexes of HE-BERT's Subclass Models on Tea and Snacks Dataset.

	Tea				Snacks			
BERT model	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1
Subclass 1	99.42%	99.49%	99.38%	99.44%	98.91%	98.92%	98.76%	98.84%
Subclass 2	97.29%	80.85%	80.99%	80.91%	96.28%	96.30%	96.05%	96.16%
Subclass 3	97.63%	97.67%	97.64%	97.65%	96.90%	96.89%	96.89%	96.89%
Subclass 4	99.43%	99.54%	99.26%	99.40%	99.37%	99.34%	99.34%	99.34%
Subclass 5	98.17%	97.28%	98.02%	97.63%	93.36%	92.69%	91.90%	92.15%
Subclass 6	97.39%	97.60%	97.56%	97.54%	94.62%	63.01%	63.46%	63.19%
Average	98.22%	95.41%	95.48%	95.43%	96.57%	91.19%	91.07%	91.10%

https://doi.org/10.1371/journal.pone.0316550.t006

superior performance. For recall, the figures are 90.86% for SVM, 90.17% for CNN, 83.71% for BAYES, and 95.48% for HS-BERT, where HS-BERT demonstrated higher recall than the other models. Lastly, considering the F1 score, the results are 91.05% for SVM, 90.20% for CNN, 84.66% for BAYES, and 95.43% for HS-BERT. In all four evaluation metrics—accuracy, precision, recall, and F1 score—HS-BERT surpassed the other models.

On the Snack dataset, the accuracy for SVM, CNN, BAYES, and HS-BERT are 81.09%, 76.16%, 73.78%, and 96.57%, respectively. HS-BERT demonstrates the highest accuracy among the models. When it comes to precision, the rates are 82.43% for SVM, 75.56% for

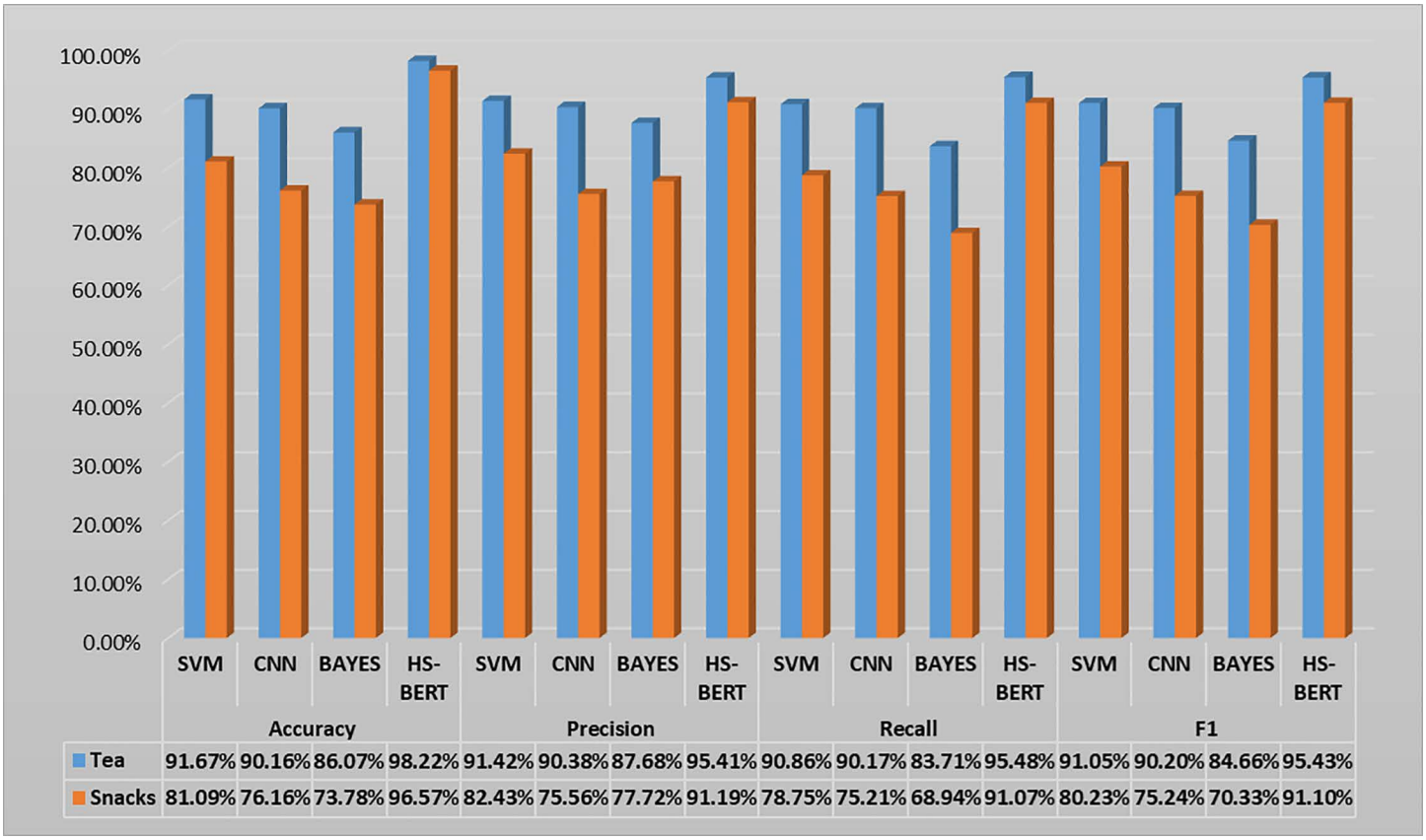


Fig 2. Comparison of HS-BERT and Other Text Classification Methods.

https://doi.org/10.1371/journal.pone.0316550.g002

CNN, 77.72% for BAYES, and 91.19% for HS-BERT, with HS-BERT leading in precision. For recall, the figures are 78.75% for SVM, 75.21% for CNN, 68.94% for BAYES, and 91.07% for HS-BERT, where HS-BERT shows the highest recall. In terms of the F1 score, the results are 80.23% for SVM, 75.24% for CNN, 70.33% for BAYES, and 91.10% for HS-BERT. Across all four metrics—accuracy, precision, recall, and F1 score—HS-BERT consistently outperforms the other models.

Experiments conducted on two distinct datasets have demonstrated that HS-BERT outperforms traditional text classification methods in terms of performance for live streaming e-commerce bullet comments. The role of live streaming bullet comment classification is pivotal in enhancing the efficiency of live streaming marketing, with its benefits highlighted in several key areas: 1) Information Extraction. It extracts crucial insights from the bullet screens, such as audience preferences, inquiries, or suggestions about the live content, providing real-time feedback to the anchor. 2) Trend Analysis. By categorizing the bullet screens, anchors can identify trending topics or patterns within the live stream, guiding content creation. 3) Audience Interaction. Classification of bullet screens helps identify audience feedback, enabling anchors to engage more effectively with viewers. These functionalities contribute significantly to the enhancement of live streaming marketing effectiveness.

6. Conclusions and future work

This research introduces an advanced BERT model, architected with a hierarchical framework, specifically tailored for the classification of e-commerce bullet comments characterized by their conciseness, diverse content, and substantial volume. Initially, the parent class BERT model is deployed to categorize the bullet comments into six predefined parent categories. Following this, the subclass BERT models are trained to further distinguish comments into their respective subcategories. This approach leverages BERT's sophisticated semantic understanding, synergistically enhancing it with the hierarchical structure's meticulous categorization capabilities. Empirical data indicates that the model in question markedly enhances the precision and efficiency of classification, facilitating deeper analysis of bullet comments, extracting actionable insights, and enabling targeted marketing strategies within the dynamic landscape of live streaming e-commerce.

To makes HS-BERT model persuasive and general, we plan to broaden it to encompass a range of linguistic contexts, extending beyond English to include French, Spanish, and numerous other languages. However, it is essential to recognize the challenges faced by HS-BERT, especially when considering the differences between alphabetic and character-based languages. For alphabetic languages such as English and Spanish, issues primarily revolve around the complexity of grammar, rich morphology, polysemy, and spelling variations. In the case of character-based languages like Japanese and Korean, challenges include the absence of clear word boundaries, difficulties in word segmentation, smaller corpus sizes, larger character sets, heavy contextual dependence, and a multitude of homophones and homographs. Addressing these language-specific hurdles is crucial for enhancing the model's effectiveness across diverse linguistic contexts. Furthermore, to enhance the generalization ability of HS-BERT model, we will test it on the datasets of e-commerce product reviews and social media reviews.

Additionally, within the sphere of live-streaming e-commerce, a troubling pattern of counterfeit viewer reviews has emerged. Influencers and businesses often hire 'professional reviewers' to post inauthentic positive reviews or inundate live streams with comments to artificially boost a product's exposure and perceived favorability. This tactic fosters a false impression of superior product quality and high sales, deceiving potential buyers. These reviewers coordinate a significant number of individuals to carry out fraudulent order placements and provide interactive engagement during live streams, all in service of creating a deceptively positive

image for the brand. Additionally, they utilize cloud control systems that enable a single phone to control multiple devices simultaneously for posting reviews, further manipulating the data to their advantage. The proliferation of inauthentic comments poses a significant threat to the understanding of viewer preference. As a result, the development of robust methods to detect and eradicate these deceptive reviews has become a pressing issue.

Supporting information

S1 Data. Supporting information data.
(ZIP)

Author contributions

Software: Huafeng Kong.

Supervision: Huafeng Kong.

Writing – original draft: Rong Zhou.

Writing – review & editing: Qing Shen.

References

1. Meng L (Monroe), Duan S, Zhao Y, Lü K, Chen S. The impact of online celebrity in livestreaming e-commerce on purchase intention from the perspective of emotional contagion. J Retail Consum Serv. 2021;63:102733. https://doi.org/10.1016/j.jretconser.2021.102733
2. Zhou R, Tong L. A Study on the influencing factors of consumers' purchase intention during livestreaming e-commerce: the mediating effect of emotion. Front Psychol. 2022;13:903023. https://doi.org/10.3389/fpsyg.2022.903023 PMID: 35615168
3. Shen Q, Wen Y han, Comite U. E-commerce live streaming danmaku classification through LDA-enhanced BERT-TextCNN model. International Journal of Information Technologies and Systems Approach (IJITSA). 2024;17(1):1–23. https://doi.org/10.4018/ijitsa.350301
4. Zhang Y, Li K, Qian C, Li X, Yuan Q. How real-time interaction and sentiment influence online sales? Understanding the role of live streaming danmaku. J Retail Consum Serv. 2024;78:103793. https://doi.org/10.1016/j.jretconser.2024.103793
5. Zeng Q, Guo Q, Zhuang W, Zhang Y, Fan W. Do real-time reviews matter? Examining how bullet screen influences consumers' purchase intention in live streaming commerce. Inf Syst Front. 2022;25(5):2051–67. https://doi.org/10.1007/s10796-022-10356-4
6. Zhang M, Ma X, Chen H. It pays to diversify: effect of bullet-screen comment diversity on payment. Psychology & Marketing. 2024;n/a(n/a).
7. Reinikainen H, Munnukka J, Maity D, Luoma-aho V. 'You really are a great big sister' – parasocial relationships, credibility, and the moderating role of audience comments in influencer marketing. Journal of Marketing Management. 2020;36(3–4):279–98. https://doi.org/10.1080/0267257x.2019.1708781
8. Dragoni M, Federici M, Rexha A. An unsupervised aspect extraction strategy for monitoring real-time reviews stream. Information Processing & Management. 2019;56(3):1103–18. https://doi.org/10.1016/j.ipm.2018.04.010
9. Li W, Jin B, Quan Y. Review of research on text sentiment analysis based on deep learning. Open Access Library Journal. 2020;7(3):1–8.
10. Zhao B, Jin W, Del Ser J, Yang G. ChatAgri: exploring potentials of ChatGPT on cross-linguistic agricultural text classification. Neurocomputing. 2023;557:126708. https://doi.org/10.1016/j.neucom.2023.126708
11. Cao Y, Sun Z, Li L, Mo W. A study of sentiment analysis algorithms for agricultural product reviews based on improved BERT model. Symmetry. 2022;14(8):1604. https://doi.org/10.3390/sym14081604
12. Su M, Cheng D, Xu Y, Weng F. An improved BERT method for the evolution of network public opinion of major infectious diseases: case study of COVID-19. Expert Systems with Applications. 2023;233:120938. https://doi.org/10.1016/j.eswa.2023.120938
13. Ma Y, Liu X, Zhao L, Liang Y, Zhang P, Jin B. Hybrid embedding-based text representation for hierarchical multi-label text classification. Expert Systems with Applications. 2022;187:115905. https://doi.org/10.1016/j.eswa.2021.115905

14. Cerri R, Barros R, Carvalho A. A genetic algorithm for hierarchical multi-label classification. Proceedings of the 27th Annual ACM Symposium on Applied Computing. 2012:250–5.
15. Huang W, Chen E, Liu Q, Chen Y, Huang Z, Liu Y. Hierarchical multi-label text classification: an attention-based recurrent network approach. Proceedings of the 28th ACM International Conference on Information and Knowledge Management. 2019:1051–60.
16. Jin W, Zhao B, Zhang Y, Huang J, Yu H. WordTransABSA: enhancing aspect-based sentiment analysis with masked language modeling for affective token prediction. Expert Systems with Applications. 2024;238:122289. https://doi.org/10.1016/j.eswa.2023.122289
17. Lin H-CK, Wang T-H, Lin G-C, Cheng S-C, Chen H-R, Huang Y-M. Applying sentiment analysis to automatically classify consumer comments concerning marketing 4Cs aspects. Applied Soft Computing. 2020;97:106755. https://doi.org/10.1016/j.asoc.2020.106755
18. Devlin J, Chang M-W, Lee K, Toutanova K. Bert: pre-training of deep bidirectional transformers for language understanding. arXiv preprint. 2018;1810.04805. https://doi.org/10.48550/arXiv.1810.04805
19. Jin D, Jin Z, Zhou JT, Szolovits P. Is BERT really robust? A strong baseline for natural language attack on text classification and entailment. AAAI. 2020;34(05):8018–25. https://doi.org/10.1609/aaai.v34i05.6311
20. Kong J, Wang J, Zhang X. Hierarchical BERT with an adaptive fine-tuning strategy for document classification. Knowledge-Based Systems. 2022;238:107872. https://doi.org/10.1016/j.knosys.2021.107872
21. Maraj A, Martin M, Makrehchi M, editors. A more effective sentence-wise text segmentation approach using BERT. Document Analysis and Recognition–ICDAR 2021: 16th International Conference. 2021;16:16.
22. Su J, Dai Q, Guerin F, Zhou M. BERT-hLSTMs: BERT and hierarchical LSTMs for visual storytelling. Computer Speech & Language. 2021;67:101169. https://doi.org/10.1016/j.csl.2020.101169
23. Araci D. Financial sentiment analysis with pre-trained language models. arXiv preprint. 2019.
24. Wang C, Jiang H, Chen T, Liu J, Wang M, Jiang S, et al. Entity understanding with hierarchical graph learning for enhanced text classification. Knowledge-Based Systems. 2022;244:108576. https://doi.org/10.1016/j.knosys.2022.108576
25. Stein RA, Jaques PA, Valiati JF. An analysis of hierarchical text classification using word embeddings. Information Sciences. 2019;471:216–32. https://doi.org/10.1016/j.ins.2018.09.001
26. Peng H, Li J, He Y, Liu Y, Bao M, Wang L, et al. Large-scale hierarchical text classification with recursively regularized deep graph-CNN. Proceedings of the 2018 World Wide Web Conference. 2018. p. 1063–72.
27. Meng Y, Shen J, Zhang C, Han J. Weakly-supervised hierarchical text classification. AAAI. 2019;33(01):6826–33. https://doi.org/10.1609/aaai.v33i01.33016826
28. Mao Y, Tian J, Han J, Ren X. Hierarchical text classification with reinforced label assignment. arXiv preprint. 2019. https://doi.org/arXiv:1908.10419
29. Kowsari K, Brown D, Heidarysafa M, Meimandi K, Gerber M, Barnes L, editors. HDLTex: hierarchical deep learning for text classification. 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA). 2017:18–21.
30. Jiang T, Wang D, Sun L, Chen Z, Zhuang F, Yang Q. Exploiting global and local hierarchies for hierarchical text classification. arXiv preprint. 2022. https://doi.org/arXiv:2205.02613
31. Gargiulo F, Silvestri S, Ciampi M, De Pietro G. Deep neural network for hierarchical extreme multi-label text classification. Applied Soft Computing. 2019;79:125–38. https://doi.org/10.1016/j.asoc.2019.03.041
32. Wang Z, Wang P, Huang L, Sun X, Wang H. Incorporating hierarchy into text encoder: a contrastive learning approach for hierarchical text classification. arXiv preprint. 2022. https://doi.org/10.48550/arXiv.2203.03825
33. Rojas K, Bustamante G, Oncevay A, Cabezudo M. Efficient strategies for hierarchical text classification: external knowledge and auxiliary tasks. arXiv preprint. 2020. https://doi.org/arXiv:2005.02473
34. Montag C, Yang H, Elhai JD. On the psychology of TikTok use: a first glimpse from empirical findings. Frontiers in Public Health. 2021;9.
35. Chen H, Dou Y, Xiao Y. Understanding the role of live streamers in live-streaming e-commerce. Electronic Commerce Research and Applications. 2023;59:101266. https://doi.org/10.1016/j.elerap.2023.101266
36. Zhong Y, Zhang Y, Luo M, Wei J, Liao S, Tan K-L, et al. I give discounts, I share information, I interact with viewers: a predictive analysis on factors enhancing college students' purchase intention in a live-streaming shopping environment. Young Consumers. 2022;23(3):449–67. https://doi.org/10.1108/yc-08-2021-1367

37. Construction of evaluation indicators for the quality of logistics services on live e-commerce platforms. AJBM. 2023;5(15). https://doi.org/10.25236/ajbm.2023.051525

38. Xu W, Cao Y, Chen R. A multimodal analytics framework for product sales prediction with the reputation of anchors in live streaming e-commerce. Decision Support Systems. 2023:114104.

39. Farivar S, Wang F, Turel O. Followers' problematic engagement with influencers on social media: An attachment theory perspective. Computers in Human Behavior. 2022;133:107288. https://doi.org/10.1016/j.chb.2022.107288

40. Jun S, Yi J. What makes followers loyal? The role of influencer interactivity in building influencer brand equity. JPBM. 2020;29(6):803–14. https://doi.org/10.1108/jpbm-02-2019-2280

41. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. Pytorch: an imperative style, high-performance deep learning library. Advances in Neural Information Processing Systems. 2019;32.

42. Huang XS, Perez F, Ba J, Volkovs M. Improving Transformer Optimization Through Better Initialization. In: Hal D III, Aarti S, editors. Proceedings of the 37th International Conference on Machine Learning; Proceedings of Machine Learning Research: PMLR; 2020. p. 4475–83.

43. Guan L. Weight prediction boosts the convergence of adamw. InPacific-Asia Conference on Knowledge Discovery and Data Mining 2023 May 25 (pp. 329-340). Cham: Springer Nature Switzerland.