

Adaptive BERT for Semantic Alignment and Intelligent Triage of Multilingual Inquiries in Cross-Border E-Commerce

Danyan Chen

Associate Professor, School of Digital Commerce, Wuxi Vocational Institute of Commerce, Wuxi, 214153, China,
E-mail: DanyanChendy@outlook.com

Project Management

Received March 4, 2026; revised April 30, 2026; July 14, 2026; accepted September 9, 2026
Available online September 20, 2026

Abstract: Currently, in the field of cross-border e-commerce, there are problems such as insufficient adaptation of general pre-training models to domain features, significant differences in the semantic gap between English-Chinese cross-language and English-Russian scenarios, low training efficiency and high deployment costs when implementing complex models. In this regard, the study proposes a domain-adaptive Multilingual Bidirectional Encoder Representations from Transformers (MBERT) model that integrates a Domain Adaptation Layer (DAL) and a Cross-Language Alignment Module (CLAM). This study first performs domain adaptation optimization through pre-training a Domain-Distinguish Task (DDT) BERT model. On this basis, MBERT is used to provide the model with a cross-language semantic representation, and the DDT mechanism and Retrieval-Augmented Generation (RAG) technology are finally integrated. The experimental results showed that, on the SemEval-2016 attribute-level emotion annotation data set, the proposed model achieved average accuracies of 82.2%, 83.2%, and 88.0% in the English-Chinese, English-Russian, and English-Spanish scenarios, respectively, demonstrating the best performance. Pre-training took 6.3 h and converged in 16 rounds. Inference took 19ms/sentence and occupied 9GB of video memory. Adding DAL alone increased the F1 value in the English-Spanish scene by 4.1%, and adding CLAM alone increased the English-Chinese accuracy by 4.2%.

Keywords: Domain adaptation; bidirectional encoder representations from transformers (BERT); multi-language; semantic alignment; cross-border e-commerce.

Copyright © Journal of Engineering, Project, and Production Management (EPPM-Journal).
DOI 10.32738/JEPPM-2026-298

1. Introduction

Amid the evolving global digital economy, Cross-Border E-Commerce (CBEC) has emerged as a core driver of innovation in international trade, with expanding market scale and service coverage (Chen et al., 2024). iResearch's 2024 report indicates China's cross-border export e-commerce achieved a 17.0% CAGR, accounting for 35.3% of total merchandise exports, demonstrating strong adaptability to global supply chain restructuring and online consumption trends (Adachi and Ogawa, 2025). User inquiry serves as a critical nexus between consumer needs and platform responses, where the accuracy of semantic analysis and triage efficiency directly impact satisfaction and conversion rates (Yan, 2024). However, multilingual semantic heterogeneity, domain-scenario ambiguity, and structural limitations of traditional models constitute core bottlenecks hindering CBEC service intelligence (Goud and Garg, 2025). Current CBEC Multi-Lingual Inquiry (MLI) processing predominantly relies on manual translation and triage, facing three key constraints (Noorian et al., 2024): processing delays that fail to meet real-time transaction demands, low triage accuracy that leads to resource misallocation, and insufficient depth of semantic analysis. General machine translation and understanding models lack domain adaptation and struggle to capture CBEC-specific semantic information (Ma et al., 2024).

To overcome the above limitations, academia and industry have begun exploring solutions based on pre-trained language models. Among them, the Bidirectional Encoder Representations from Transformers (BERT) model has become a research hotspot in CBEC for Natural Language Processing (NLP) tasks, owing to the advantages of its bidirectional attention mechanism in semantic representation. The core value of the BERT architecture lies in capturing multilingual semantic associations through bidirectional context modeling, effectively addressing problems such as polysemy and semantic ambiguity in cross-border scenarios. Its pre-training-fine-tuning paradigm can quickly adapt to specific tasks such as CBEC inquiry processing and semantic alignment, significantly reducing the cost of domain data annotation. In existing research,

the application of BERT and its derivative models in the e-commerce field has shown initial value (Zou et al., 2024). Gardazi et al. noted in their review that the BERT model outperforms traditional deep learning models in tasks such as e-commerce review analysis and product recommendation, owing to its deep semantic encoding capabilities (Gardazi et al., 2025). Karabila et al. proposed the BERT-enhanced sentiment analysis model. By fine-tuning the corpus in the e-commerce review domain, the model achieved accurate identification of emotional tendencies in personalized recommendation scenarios, confirming the key role of domain adaptation in improving the performance of the BERT model (Karabila et al., 2024). Zhao et al. effectively mined potential needs from user reviews for the e-commerce recommendation task by integrating BERT semantic features with deep neural networks, further demonstrating the feasibility of integrating BERT with domain tasks (Zhao et al., 2024). Bagiyalakshmi and A. C. Kothandaraman optimized the deep learning model architecture and applied the BERT semantic representation to e-commerce website sentiment analysis, thereby significantly improving the accuracy of consumer experience evaluation (Bagiyalakshmi and Kothandaraman, 2024).

However, existing research still has obvious shortcomings. Although the general BERT model demonstrates excellent semantic representation capabilities in NLP, its versatility is limited by the training corpus, and its semantic alignment accuracy in the CBEC vertical field still has considerable room for improvement (Bilal and Almazroi, 2023). Therefore, this research question is whether the domain adaptation module can significantly improve the cross-language semantic alignment accuracy of CBEC multilingual inquiries. In view of this, the study proposes a CBEC-MLI semantic alignment and intelligent triage method based on domain-adaptive BERT, aiming to achieve efficient semantic matching and accurate MLI triage in CBEC scenarios. The innovation of the research lies in the integration of Multilingual BERT (MBERT), Domain-distinguish Task (DDT), and Retrieval-Augmented Generation (RAG) technology. It strengthens CBEC-specific semantic capture through domain adaptation, relies on a multilingual architecture to address semantic heterogeneity, and combines RAG to improve the accuracy of triage decisions. The research builds a complete technical-logical framework based on cross-language representation learning theory and domain self-adaptive transfer learning. The cross-language representation learning theory advocates achieving semantic space mapping alignment for different languages through shared encoders, providing a fundamental theoretical support for the MBERT architecture; domain self-adaptive transfer learning focuses on alleviating the catastrophic forgetting of general pre-trained models in vertical scenarios, corresponding to the two core improvement mechanisms of hierarchical parameter freezing and DDT domain pre-training described in the text. At the same time, the retrieval-enhanced generation theory is introduced, using external domain corpus knowledge to compensate for the semantic representation defects caused by limited labeled samples, forming a three-level progressive theoretical development framework of cross-language representation foundation, domain transfer optimization, and retrieval knowledge reinforcement.

2. Methodology

2.1. Construction of Domain-Adaptive BERT Model

In CBEC scenarios, multilingual semantic alignment and accurate triage demand robust cross-lingual understanding capabilities. From an industry perspective, expanding global trade has amplified CBEC transaction volumes, yet cultural context differences and localized industry terminology cause traditional machine translation to produce semantic deviations, resulting in over 40% inquiry processing delays or triage errors that directly impact customer satisfaction and repurchase rates. User demands also stratify significantly: B-side buyers prioritize professional terms (batch quotations, customs policies), while C-end consumers focus on lifestyle expressions (logistics, after-sales support). BERT addresses these challenges through its deep Transformer architecture and bidirectional pre-training mechanism (Hung and Huy, 2025; Arslan, 2024), enabling deep integration of general semantic knowledge with domain-specific tasks via pre-training and fine-tuning. The BERT architecture is shown in Fig. 1.

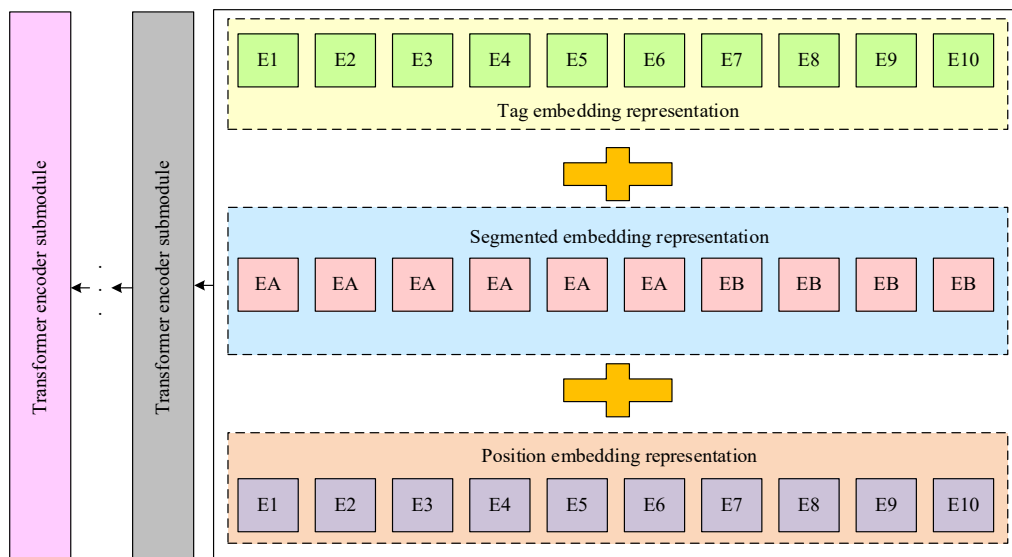


Fig. 1. BERT model structure

In Fig. 1, BERT processes input sequences of words or subwords, converted into tokens by a tokenizer that segments text into model-compatible units based on the pre-trained vocabulary (Gürbüz and Kotan, 2025). During pre-training, input sequences undergo word, segment, and position embedding to form initial vectors, then propagate through multi-layer Transformer encoders (Hao et al., 2023). The Transformer's attention mechanism is the core of semantic encoding. The query Q_i , key K_i , and value V_i of the i -th header is generated via linear projection, and the calculation is given by Eq. (1).

$$\begin{cases} Q_i = XW_i^Q \\ K_i = XW_i^K \\ V_i = XW_i^V \end{cases} \quad (1)$$

In Eq. (1), X is the input sequence. Q , K , and V are query, key, and value, and W_i^Q , K , W_i^V are the corresponding projection matrices (Li et al., 2025). On this basis, the i -th attention head calculates the feature weight by scaling the dot product attention, and the expression is as shown in Eq. (2).

$$Attention(Q_i, K_i, V_i) = Soft\ max(\frac{Q_i K_i^T}{\sqrt{d_k}}) V_i \quad (2)$$

In Eq. (2), d_k is the key and query dimension. In order to integrate multi-dimensional attention features, the model splices the outputs of all attention heads and converts them into a unified-dimensional feature vector through the output projection matrix, which is the final output of multi-head attention, as shown in Eq. (3).

$$MultiHead(Q, K, V) = Concat(head_1, head_2, \dots, head_h) W^O \quad (3)$$

In Eq. (3), W^O is the output projection matrix. Among them, the output $head_i$ expression of the i -th attention head is shown in Eq. (4).

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \quad (4)$$

There are two forms of positional encoding used in Transformer models, corresponding to even and odd indices of the embedding dimension. The expression of even-numbered positions is shown in Eq. (5).

$$PE_{(pos, 2i)} = \sin(\frac{pos}{1000^{2i/d}}) \quad (5)$$

In Eq. (5), $PE_{(pos, i)}$ is the position code, pos is the position index, and d is the embedding dimension. The odd position calculation is given by Eq. (6).

$$PE_{(pos, 2i+1)} = \cos(\frac{pos}{1000^{2i/d}}) \quad (6)$$

Each encoder layer comprises multi-head self-attention and feed-forward networks for semantic extraction and integration. BERT performs Masked Language Model (MLM) and Next Sentence Prediction (NSP) tasks on the final layer output (Zardak et al., 2023) (Koy and Çolak, 2024): MLM predicts masked tokens via classification, while NSP determines sentence continuity through binary classification. DDT enhances BERT's domain adaptation by embedding domain differentiation tasks during pre-training, guiding the model to learn cross-domain feature variations for more accurate domain-specific representations (Hicham and Nassera, 2024). The BERT-DDT architecture is illustrated in Fig. 2.

In Fig. 2, BERT-DDT achieves domain adaptation by jointly training domain and emotion classification tasks. The base architecture uses BERT's multi-layer Transformer encoder to generate token-level features and sentence-level [CLS] representations. Two parallel branches extend from this: (1) an emotion classification branch that predicts sentiment via a linear classifier on global features, serving the core task (Kassem et al., 2024); and (2) a domain classification branch incorporating a Gradient Reversal Layer (GRL) that reverses gradients during backpropagation, which ensures the model learns domain-invariant emotional features while maintaining domain discrimination capability.

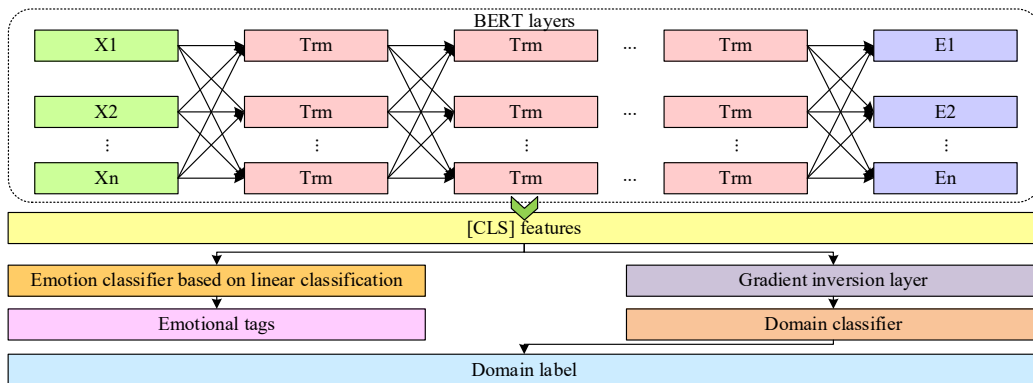


Fig. 2. Architecture diagram of cross-domain sentiment analysis based on BERT-DDT

2.2. Cross-Language Pre-Training Model Based on MBERT-DDT

Although the BERT model pre-training corpus spans multiple languages, it lacks targeted optimization for vertical fields such as CBEC, making it difficult to accurately capture cross-language semantic associations among terms. Therefore, this study introduces the MBERT model as the basic architecture. MBERT strengthens cross-language semantic representation capabilities by expanding the language coverage and domain diversity of the pre-training corpus, as shown in Fig. 3.

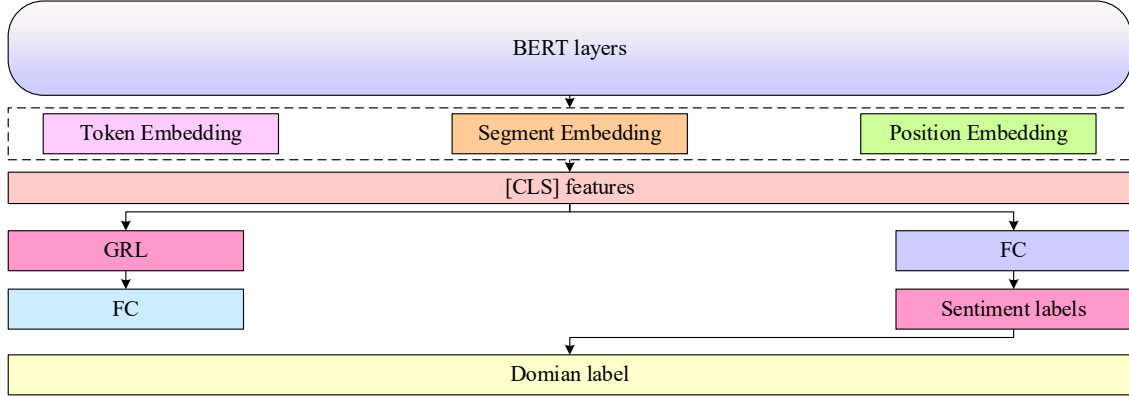


Fig. 3. MBERT model architecture

In Fig. 3, MBERT extends BERT's core architecture for multilingual scenarios through a three-layer structure: input embedding, deep encoding, and flexible output. The input layer combines word embedding (mapping multilingual vocabularies to a unified semantic space), language embedding (identifying source language), and positional embedding (preserving word order) (Kansal and Diwaker, 2025). The encoder comprises 12 Transformer layers, each containing multi-head self-attention (capturing cross-lingual associations via parallel attention weights) and a Feedforward Neural Network (FNN) for nonlinear transformation, enhancing domain-specific feature discrimination. The FNN formulation is given in Eq. (7).

$$FFN(x) = \max(0, xW_1 + b_1)W_2 + b_2 \quad (7)$$

In Eq. (7), x is the input vector. $\max$ is the activation function. W_1 and b_1 are the weight matrices and bias terms of the middle layer. W_2 and b_2 are the weight matrix and bias terms of the output layer. The output layer produces sentence-level global features via [CLS] tokens, which not only support downstream tasks such as emotion classification but also, through fine-tuning, adapt to the semantic alignment requirements of CBEC-MLI, laying an architectural foundation for the subsequent introduction of domain adaptation modules (Maiga et al., 2025). Next, the model introduces the MLM pre-training task and uses the cross-entropy loss function for training (El-Ansari and Beni-Hssane, 2023). The expression of the probability $p(x_a|\hat{x}_a)$ of the correct word from the first formula is shown in Eq. (8).

$$p(x_a|\hat{x}_a) = \text{Soft max}(W_m h_a + b_m) \quad (8)$$

In Eq. (8), $\hat{x}_a$ is the input text after random mask processing. a is the sequence position. x_a is the true label of the masked word at position a . h_a is the context vector corresponding to position a calculated by the BERT model. W_m and b_m are weight matrices and bias vectors specific to the MLM task. Softmax is the normalized exponential function. The calculation of the cross-entropy loss function L_m is shown in Eq. (9).

$$L_m = -\frac{1}{N_S} \sum_{K=1}^{N_S} \frac{1}{N_T^K} \sum_{a=1}^{N_T^K} \log p(x_a|\hat{x}_a) \quad (9)$$

In Eq. (9), $\sum_{a=1}^{N_T^K}$ is all the masked positions in the K -th sentence. $\frac{1}{N_T^K} \sum_{a=1}^{N_T^K}$ is the average loss of the K -th sentence.

$\log p(x_a|\hat{x}_a)$ is the logarithm of the probability of calculating the correct word. $\sum_{K=1}^{N_S}$ is the sum of the losses of all

samples N_S in the training batch. $\frac{1}{N_S} \sum_{K=1}^{N_S}$ is the average loss of the entire batch. For the emotion classification task, this

study uses a binary classification probability distribution, i.e., the probability of a positive or negative emotion. After Softmax processing, the expression of the prediction probability distribution y_c is shown in Eq. (10).

$$y_c = \text{Soft max}(W_c h_{[CLS]} + y_c) \quad (10)$$

In Eq. (10), $h_{[CLS]}$ is the hidden state vector corresponding to the [CLS] tag obtained from the last layer of the BERT pre-trained model. W_c is the weight matrix of the sentiment classifier. y_c is the bias term of the sentiment classifier. To minimize the prediction error of the emotion classifier, this study guides the model to optimize its parameters by measuring the discrepancy between the true label and the predicted probability distribution. The cross-entropy loss function L_c is shown in Eq. (11).

$$L_c = -\frac{1}{N_S} \sum_{i=1}^{N_S} \sum_{a=1}^{N_T^K} \hat{y}_c^{(i)}(a) \log y_c^{(i)}(a) \quad (11)$$

In Eq. (11), $\hat{y}_c^{(i)}$ and $y_c^{(i)}$ are the true labels and predicted probabilities of the i -th sample. $\hat{y}_c^{(i)}(a)$ and $y_c^{(i)}(a)$ are the values of the true label vector and the predicted probability vector of the i -th sample at the a -th position. In response to the semantic alignment requirements of CBEC multi-language scenarios, this study introduces hierarchical embedding and bidirectional attention interaction on the MBERT-DDT architecture to achieve deep correlation of multi-language semantics. The cross-language pre-training model based on MBERT-DDT is shown in Fig. 4.

In Fig. 4, token-embedded maps map cross-lingual vocabulary into a shared semantic space to ensure semantic relevance. Language embedding uses binary coding to demarcate language boundaries and prevent grammatical interference. Positional embedding encodes word order in both sequences, preserving source-language modification logic while capturing target-language syntax (Pinar, 2025). These embedding features are subsequently enhanced by the DDT module through domain-specific and general-domain classification tasks, enabling the model to anchor domain-specific semantics while filtering general noise.

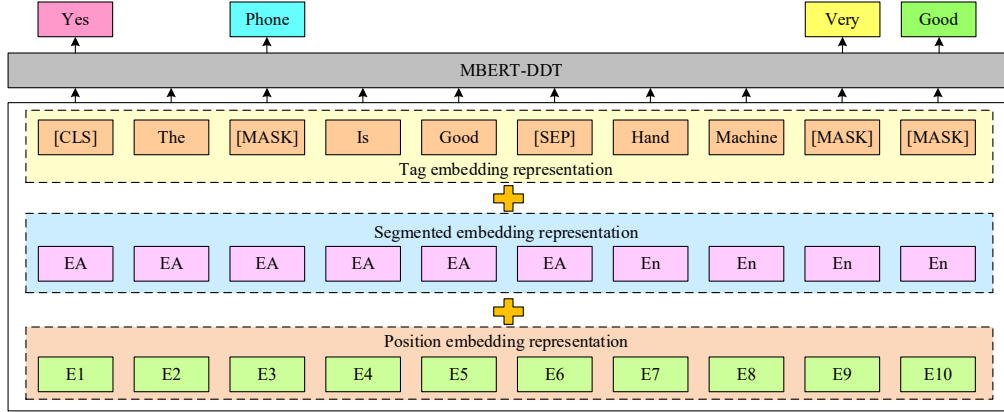


Fig. 4. Cross-language pre-training model based on MBERT-DDT

2.3. CBEC-MLI and Intelligent Triage Integrating MBERT-DDT and RAG

After the pre-training stage, the MBERT-DDT model strengthens cross-language semantic alignment between the source and target languages. It also introduces the CBEC review text corpus, which mitigates the inductive bias toward non-review texts, making it more suitable for semantic modeling in vertical fields. In view of the intelligent triage requirements of CBEC-MLI, which require rapid matching to corresponding service nodes, this study focuses on the attribute-level emotion classification task during the fine-tuning stage and further optimizes the model parameters using source-language annotation data. The architecture of the attribute sentiment classification fine-tuning model is shown in Fig. 5.

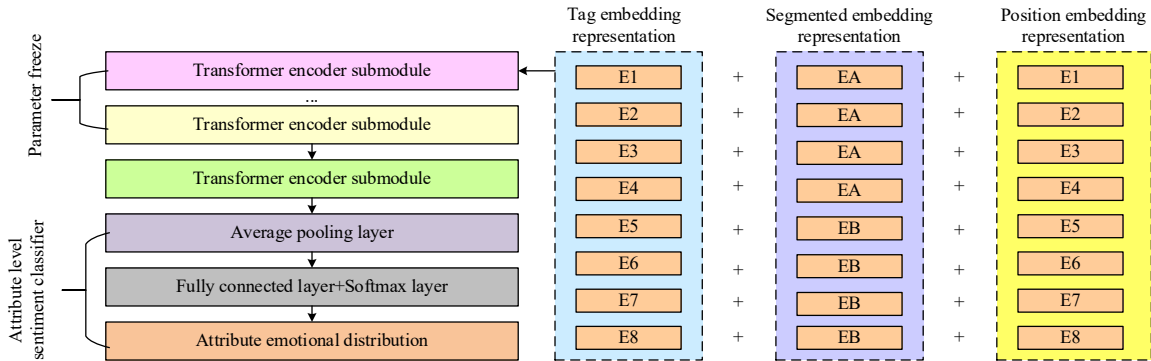


Fig. 5. Fine-tuning model for attribute emotion classification

In Fig. 5, the fine-tuning process uses the source-language annotated corpus as training data and, by constructing an input sequence in the form of $\langle \text{attribute}, \text{sentence} \rangle$, guides the model to learn the associated features of attribute words and sentence emotions, thereby achieving accurate modeling of attribute-level semantic representations. Assuming that the source language sample set in the fine-tuning stage is $X_S = \{x_1, x_2, \dots, x_n\}$. After encoding by the cross-language pre-training model, the attribute-level sentence representation H_S can be obtained, as shown in Eq. (12).

$$H_S = \text{PreT}(X_S; \delta_p) \quad (12)$$

In Eq. (12), $Pr e T$ is a cross-language pre-training model. X_S is the input fine-tuning data set, containing N constructed sequences. δ_p is the parameters of the pre-trained model. At the beginning of fine-tuning, these parameters are initialized as pre-trained weights (Teke et al., 2025). The pooled sentence vector $MeanPooling$ is shown in Eq. (13).

$$Q_S = MeanPooling(H_S) \quad (13)$$

In Eq. (13), $MeanPooling$ is the average pooling function. The output attribute emotion category probability is shown in Eq. (14).

$$y = Soft\ max(WQ_S + b) \quad (14)$$

In Eq. (14), W is the weight matrix and b is the bias vector. The fine-tuning phase uses the cross-entropy loss function as the optimization objective and updates the model parameters by measuring the discrepancy between the predicted probability distribution and the true label. The loss function L is shown in Eq. (15).

$$L = -\sum_{i=1}^U \log P(x_s^{(i)}|\delta) \quad (15)$$

In Eq. (15), $x_s^{(i)}$ is the training sample, δ is the set of unfrozen parameters and attribute emotion classifier parameters, and U is the data set. After fine-tuning, the model acquires both cross-lingual semantic alignment and attribute-level sentiment classification capabilities, enabling direct target language input during testing without additional annotation. To address deviations in semantic understanding stemming from outdated domain knowledge and poor triage accuracy in CBEC-MLI processing, this study introduces RAG technology to enable precise domain knowledge retrieval and dynamic adaptation. The retriever is fine-tuned on CBEC-specific datasets to optimize domain term matching, while a dynamic weight adjustment mechanism enables adaptive retrieval scope and knowledge fusion weighting based on domain relevance and context complexity. The RAG architecture is shown in Fig. 6.

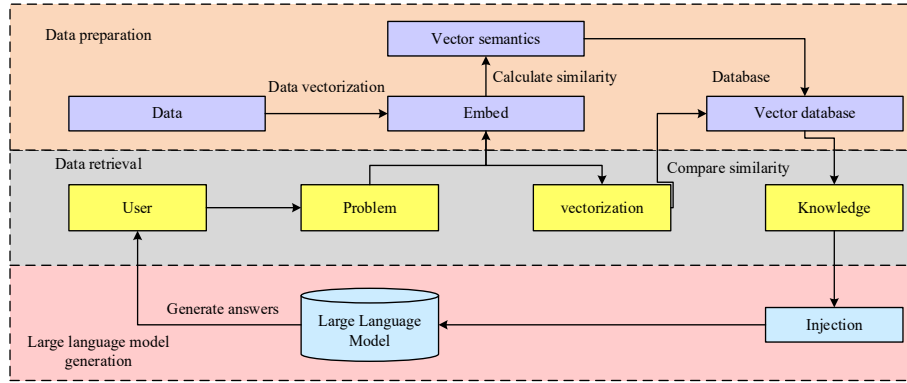


Fig. 6. RAG architecture

Fig. 6 illustrates the RAG system architecture comprising a retriever, external knowledge base, and generator. Upon receiving user input, the retriever performs semantic encoding and vector similarity calculation against the knowledge base to recall Top-K relevant documents (Coban et al., 2024). These documents are integrated with the original query to form an enhanced input sequence for the generator. During decoding, the generator leverages its language modeling capabilities to ensure grammatical correctness and multilingual fluency, while employing attention mechanisms to ground its outputs in retrieved domain knowledge. A dynamic update interface enables real-time policy ingestion and incremental updates to the vector space, ensuring the system maintains up-to-date domain knowledge.

3. Results

3.1. Experimental Setup

To ensure the experiment's generalization and adaptability, this study used two data sources: a public corpus and a real-world business corpus. The public corpus used the SemEval-2016 attribute-level emotion annotation dataset, which includes annotation specifications for multilingual emotional expressions. The real corpus was collected from public comments on Suning.com's CBEC platform, covering two core cross-border fields: catering (service category) and notebooks (3C digital category). The former corresponds to the immediate consumption scenario, and the latter corresponds to the high customer unit price durable consumer goods scenario. The semantic features of the inquiry orders across the two scenarios differ significantly, enabling comprehensive testing of the model's domain adaptation ability. In the catering field, English (the lingua franca of global trade) was selected as the source language, and Spanish (the main consumer market in South America) and Russian (the core market in Eastern Europe) were selected as the target languages. In the notebook field, English was selected as the source language, and Chinese (China's local market) as the target language. Corpus processing follows a two-stage logic of pre-training and fine-tuning. In the pre-training stage, Google Translate API was used to construct <source language, target translation> sentence pairs to achieve multi-language semantic alignment. In the fine-tuning stage, only an annotated source-language corpus was used, and the development and training sets were split 8:2. Table 1 shows the specific configuration parameters.

The size, distribution characteristics, and category distribution of the experimental corpus directly affect the adequacy of model training and the objectivity of evaluation. The specific statistical information of the corpus used is shown in Table 2.

Table 1. Experimental configurations

Configuration category	Name	Model/Parameter
Hardware	CPU	Intel Xeon Silver 4314 @ 2.40GHz
	GPU	NVIDIA A100
	memory	256GB DDR4 3200MHz
	storage	2TB SSD + 10TB HDD
Software	operating system	CentOS 7.9
	Deep Learning Framework	TensorFlow 2.10 + Hugging Face Transformers 4.25
	translation tool	Google Translate API V3
	evaluation tool	Scikit-learn 1.2.0
Model hyperparameters	pre-training	Batch size 64, learning rate 1e-4, 15 iterations
	fine-tuning	Batch size 32, learning rate 2e-5, 25 iterations
	Domain adaptive module	dropout=0.3, Attention weight decay=0.01

Table 2. Details of corpus statistics

/	Notebook		Catering			Total
	EN (Source language)	EN-CH	EN (Source language)	EN-RU	EN-SP	
Number of training sets (pieces)	1882	1882	1925	1925	1925	9539
Number of development sets (pieces)	471	471	482	482	482	2388
Average sentence length (number of words)	13	24	14	12	15	/
Maximum sentence length (number of words)	73	132	57	64	79	/
Distribution of emotional categories (positive: negative)	1430:923	1430:923	1657:749	1657:749	1657:749	7384:3850

In Table 2, the two types of domain corpora are consistent in terms of quantity, length characteristics and emotional distribution: the size of the single-language training set was approximately 1800 to 1900 items, ensuring adequate model training. The text length increased significantly after translation (for example, the average length of the notebook EN-CH increased from 13 words to 24 words), which is consistent with differences in language expression habits between Chinese and English. The ratio of positive to negative emotions in each language direction was approximately 1.5:1 to mitigate data bias during model training. A total of 9,539 training data points and 2,388 development data points could fully support model performance evaluation in cross-language scenarios.

3.2. Domain-Adaptive BERT Model Performance Test Results

To verify the emotional feature discrimination ability and domain transfer performance of the proposed model in cross-domain scenarios, this study inputs the text data from the source and target domains into the model, respectively and extracts the 768-dimensional feature vector for each text encoded by the model. It then uses the t-SNE algorithm to reduce the high-dimensional features to 2D for visualization, as shown in Fig. 7.

Fig. 7(a) shows a source domain visualization: feature vectors exhibit clear separation by sentiment polarity, with positive samples concentrated on the left and negative samples on the right; minimal overlap indicates effective capture of emotional features and distinguishable semantic encoding. Fig. 7(b) presents target domain visualization with similarly distinct positive/negative clustering, showing only minor edge intersections. To balance cross-lingual general knowledge retention with domain-specific learning while mitigating catastrophic forgetting, this study employs hierarchical parameter freezing during fine-tuning. All experiments were independently repeated three times. The variance intervals for each evaluation index were measured and reported as 0.8%-1.5%, and the 95% confidence intervals were $\pm 1.2\%$. This precisely quantified the range of fluctuations in the model performance indicators, effectively avoiding interference from random errors in a single experiment on the conclusion, and significantly improving the reproducibility and credibility of the experimental results. Systematic exploration of freezing layer effects across language directions and domains is shown in Fig. 8.

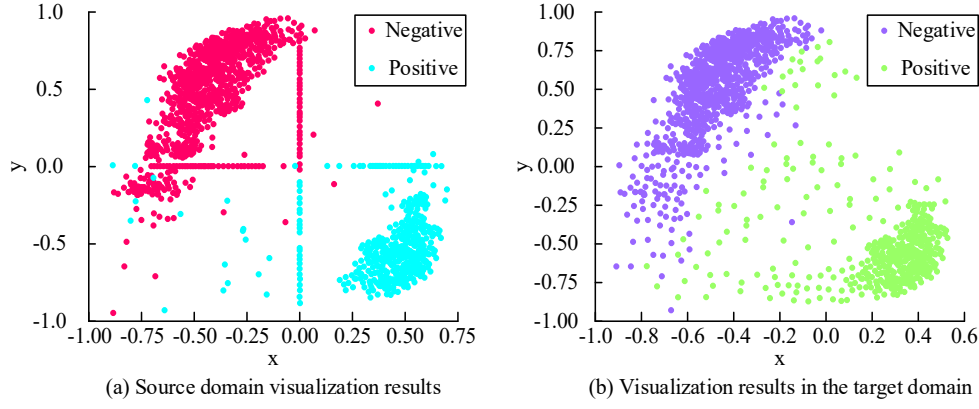


Fig. 7. Visualization of model features

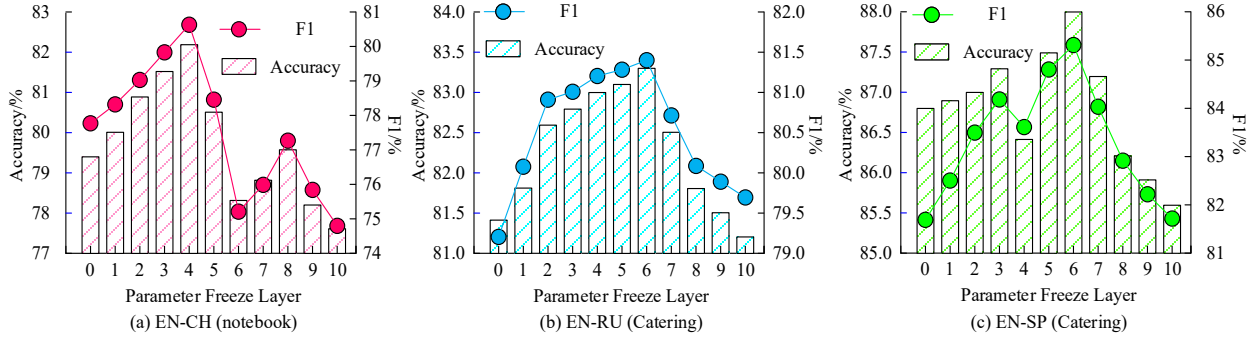


Fig. 8. Comparison of the effects of freezing layers on parameters

Fig. 8(a) examines layer freezing effects in EN-CH scenarios, with optimal performance at 4 frozen layers (82.2% accuracy, 80.6% F1). Fig. 8(b) shows EN-RU results, where the optimum shifts to 6 layers (83.3% accuracy, 81.4% F1). Fig. 8(c) presents EN-SP results, peaking at 6 frozen layers with 88.0% accuracy and 85.3% F1, the highest across all scenarios. The F1 score improvement from 84.8% (layer 5) to 85.3% (layer 6) indicates that accurate attribute-word distinction requires mid-level parameter optimization. Five mainstream CBEC multilingual models were compared: MT-DNN (A), MBERT (B), ERNIE-M (C), SSL-MBERT (D), and XLM-RoBERTa (E). Their accuracy across language directions is shown in Fig. 9.

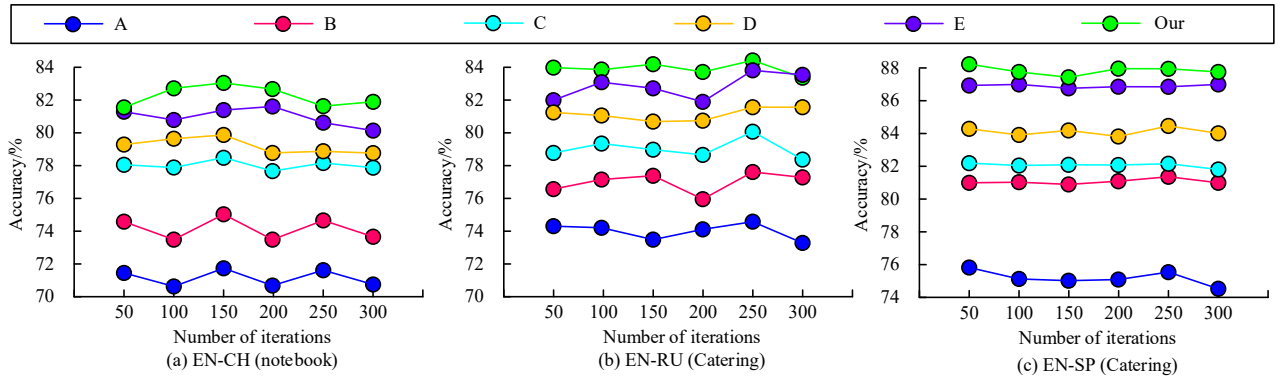


Fig. 9. Comparison of cross-language classification accuracy of various models

Fig. 9(a) presents cross-lingual accuracy for English-Chinese (notebook domain): MT-DNN (71%), MBERT (74%), and the proposed model (82%). Fig. 9(b) shows English-Russian (catering) results: SSL-MBERT achieved 81% with unlabeled Russian data, XLM-RoBERTa(E) reached 83%, while the proposed model attained 84%; its domain-adaptive module effectively extracted core semantics (taste, service, delivery) and mitigated Russian morphological interference. Fig. 9(c) displays English-Spanish (catering) performance, with all models exceeding 74% and the proposed model reaching 88%.

3.3. Application Effect of Domain-Adaptive BERT Model in CBEC

To comprehensively verify the engineering value and performance advantages of the proposed model, this study conducted comparative experiments from three dimensions: training efficiency, deployment adaptability, and module component effectiveness. The system evaluated its practical value in CBEC multi-language scenarios. Training efficiency was measured

by two indicators: pre-training time and convergence rounds. The comparison results of different models are shown in Fig. 11.

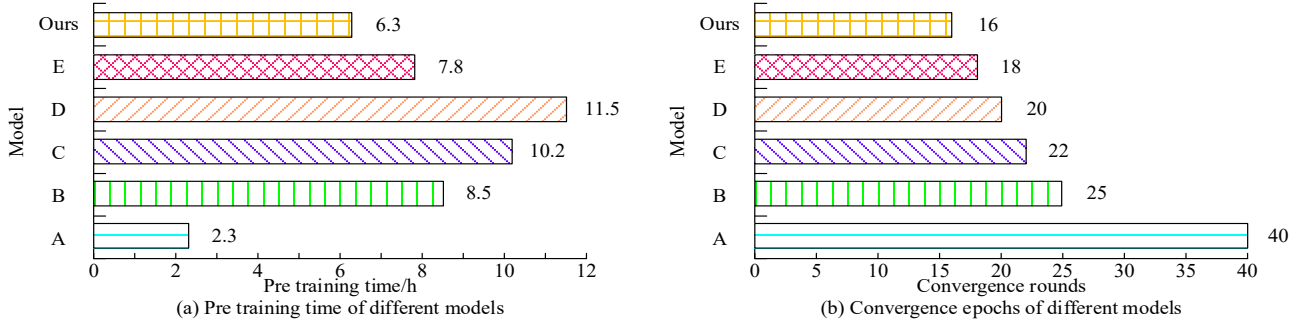


Fig. 11. Comparison of pre-training time and convergence epochs for different models

Fig. 11(a) compares pre-training times: MT-DAN required only 2.3 hours due to its shallow architecture and the absence of large-scale multilingual pre-training; SSL-MBERT took the longest at 11.5 hours due to pseudo-label generation and iterative optimization on unlabeled data; the proposed model required 6.3 hours, 25.9% less than MBERT. Fig. 11(b) shows the convergence rounds for MT-DAN (40), MBERT (25), and XLM-RoBERTa (18), with the proposed model achieving the fastest convergence at 16 rounds. To assess deployment feasibility, model parameters, inference time, and GPU memory usage are compared across six models, as shown in Fig. 12.

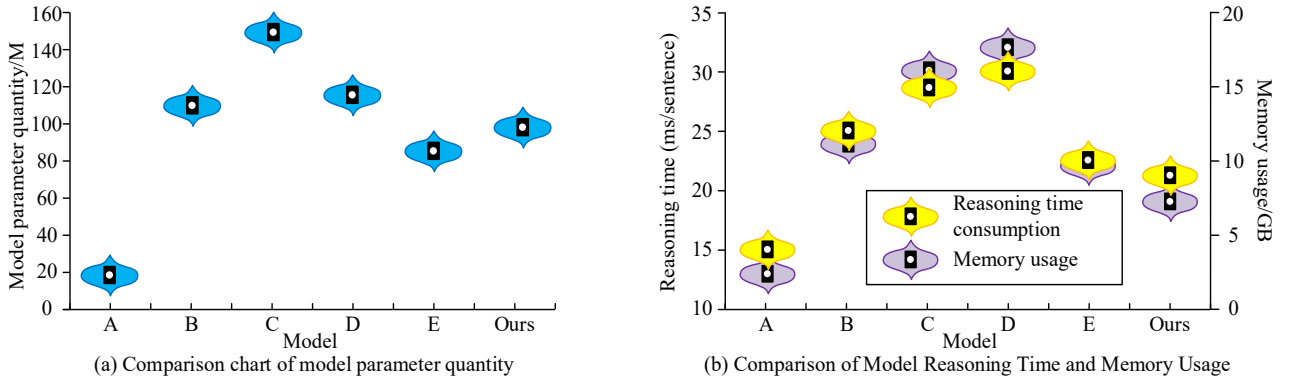


Fig. 12. Comparison of different model parameter quantities, inference time, and existing comparisons

Fig. 12(a) compares model parameters: MT-DAN (18M), ERNIE-M (150M), SSL-MBERT (slightly higher than mBERT due to pseudo-label prediction heads), XLM-RoBERTa (85M via streamlined pre-training), and the proposed model (98M). Fig. 12(b) shows inference time and memory usage: MT-DAN was fastest at 13ms/sentence, and SSL-MBERT was slowest at 32ms/sentence; the proposed model achieved 19ms/sentence, outperforming MBERT and XLM-RoBERTa, with only 9GB of GPU memory versus XLM-RoBERTa's 10GB. To systematically analyze the independent gains of the DAL, CLAM, and RAG components as well as the collaborative effects of different combinations, the study uniformly used MBERT as the standard baseline. All experimental groups shared exactly the same dataset and hardware environment. The pre-training learning rate was traversed through $\{5e-5, 1e-4, 2e-4\}$, and the fine-tuning learning rate was traversed through $\{1e-5, 2e-5, 5e-5\}$. The dropout values were 0.1/0.3/0.5, and the number of layers for layer-wise freezing was varied across 0/4/6/8 layers. The experiments were set up with single-, dual-, and triple-module combinations, totaling 6 control groups. This precisely isolated the individual effects and interaction effects of each module. The ablation experiment results are shown in Table 3.

Table 3 Comparison of ablation experiment results

Model	EN-CH		EN-RU		EN-SP	
	Acc(%)	F1(%)	Acc(%)	F1(%)	Acc(%)	F1(%)
MBERT	74.7	73.4	77.0	72.8	81.6	75.3
MBERT+DAL	76.8	74.2	78.5	75.8	83.9	79.4
MBERT+CLAM	78.9	75.7	80.1	76.9	84.5	80.2
MBERT+RAG	77.5	74.6	78.2	74.5	82.7	77.1
MBERT+DAL+CLAM	80.1	78.2	81.5	79.4	86.2	82.6
Ours	82.2	80.6	83.2	81.3	88.0	85.3

From the stratified ablation results in Table 3, it can be seen that in the single-module scenario, DAL is good at capturing the domain-specific semantics in the cross-border e-commerce field, and the F1 score in the same language family EN-SP scenario increased by 4.1%. CLAM focuses on eliminating semantic differences across languages, and the accuracy rate in the cross-language family EN-CH increased the most (4.2%). RAG relies on the domain knowledge base to supplement scarce industry terms and integrates the retrieved semantic vectors into the encoding end to participate in prediction, thereby slightly correcting the semantic defects of the general pre-trained model. The combination of the dual-module DAL + CLAM can simultaneously address domain feature extraction and cross-language mapping, and its performance is further improved compared to a single module. However, without external knowledge reinforcement from RAG, the indicators still show obvious shortcomings. The complete integration model of the three modules achieves the best performance, demonstrating that DAL, CLAM, and RAG each have distinct responsibilities and strong synergistic benefits, and that none can be omitted.

4. Discussion and Interpretation

Feature visualization reveals clear clustering of positive/negative sentiments across source and target domains, confirming DAL's effectiveness in extracting scenario-specific features (delivery timeliness, hardware parameters) and CLAM's unified semantic space mapping, aligning with Peng et al.'s findings on domain adaptation improving vertical scene accuracy (Peng et al., 2025). Parameter freezing experiments show optimal levels at 4 layers for cross-lingual (EN-CH) and 6 layers for same-language-family (EN-RU, EN-SP) scenarios, balancing general and domain knowledge while preventing catastrophic forgetting, consistent with Kumar et al.'s hierarchical optimization logic (Kumar et al., 2025). Performance comparisons demonstrate significant advantages: average accuracies of 82.2% (EN-CH), 83.2% (EN-RU), and 88.0% (EN-SP), ranking first among the models compared. Cross-lingual alignment ratios remain stable at 90–95% with only 5% fluctuation, substantially outperforming benchmarks such as MT-DNN and addressing multilingual alignment limitations in Karn et al.'s e-commerce recommendation research (Karn et al., 2023). Ablation experiments confirm the necessity of the dual-module synergy mechanism. When only DAL was added, the F1 value of the British-Spanish scene increased by 4.1%. When only CLAM was added, the accuracy of the English-Chinese cross-language system increased by 4.2%.

Engineering indicators demonstrate greater practical value. The pre-training process takes 6.3 hours, and convergence can be achieved after 16 rounds. The inference process takes 19ms per sentence and occupies 9GB of memory, meeting the deployment requirements of mid-range GPUs. It can be seen that the model's automatic semantic parsing and triage capabilities can significantly shorten the customer service ticket processing chain, effectively reducing manual time and labor costs; at the same time, the standardized intelligent triage mechanism optimizes the customer service workflow, avoiding manual triage mismatches and omissions, and improving the standardization and efficiency of cross-border inquiry processing. Across different CBEC platforms, significant differences in business logic, user inquiry habits and the semantic expressions of each platform mean the model has not been tested across platforms. As a result, its semantic alignment accuracy and triage efficiency may fluctuate, making it difficult to ensure compatibility on non-Suning platforms. For niche languages with fewer resources, due to the scarcity of training corpus, the model's semantic representation ability is insufficient, semantic misjudgments are prone to occur during runtime, and it is difficult to accurately capture the order inquiry intention of the niche language. Under different query structures, differences in sentence length and in the expression of intent will affect the model's parsing performance. Complex sentence patterns or non-standard query structures may lead to deviations in model semantic extraction and reduce triage accuracy.

5. Conclusion

This study proposed a domain-adaptive BERT model that integrates DAL and CLAM and verified its effectiveness and practical value through systematic experiments. The study found that the dual-module collaboration mechanism effectively solved the core problem, that DAL accurately extracted scene-based features, and that CLAM realized unified semantic mapping across different language affinities. The combination of the two raised the model's average accuracy to 82.2%, 83.2% and 88.0% in the English-Chinese, English-Russian and English-Spanish scenarios, which outperformed MT-DNN, MBERT and other models. The hierarchical parameter freezing strategy optimized the knowledge balance. Experimental results showed that cross-language (English-Chinese) optimal freezing was 4 layers, whereas the same-language family (English-Russian, English-Spanish) was optimally frozen at 6 layers. This confirmed that this strategy can avoid catastrophic forgetting. Model pre-training took 6.3 hours and 16 rounds of convergence. Inference took 19ms/sentence and occupied 9GB of video memory. It was suitable for mid-range GPU deployment and addressed the high implementation costs of traditional models. The model proposed in the study needs to improve its accuracy in processing extreme emotions. Future research can expand multimodal inquiry-processing capabilities to better meet CBEC's needs for intelligent upgrades.

Funding

The research is supported by the 2025 Jiangsu Provincial Education and Science Planning Key Project: Practical Research on Vocational Skills Training for International Students in Higher Vocational Colleges under the Background of the Belt and Road Initiative (No.: B/2025/02/11).

Institutional Review Board Statement

Not applicable.

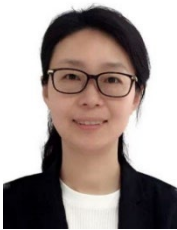
Declaration of Artificial Intelligence (AI) Tools

The author used DeepL Write solely for language editing and to improve readability. The author reviewed and verified all content and takes full responsibility for the accuracy and integrity of the manuscript.

References

- Adachi, Y. and Ogawa, H. (2025). "Cross-border shopping, E-commerce, and consumption tax revenues in Japan," *Asia-Pacific Journal of Regional Science*, 9(1), 605-627, DOI: 10.1007/s41685-025-00380-y.
- Arslan, S. (2024). "Application of BiLSTM-CRF model with different embeddings for product name extraction in unstructured Turkish text," *Neural Computing and Applications*, 36(15), 8371-8382, DOI: 10.1007/s00521-024-09532-1.
- Bagiyalakshmi, N. and Kothandaraman, A.C. (2024). "Optimized Deep Learning Model for Sentimental Analysis to Improve Consumer Experience in E-Commerce Websites," *Journal of Computing and Artificial Intelligence*, 2(3), 41-54, DOI: 10.69996/jcai.2024014.
- Bilal, M. and Almazroi, A.A. (2023). "Effectiveness of Fine-tuned BERT Model in Classification of Helpful and Unhelpful Online Customer Reviews," *Electronic Commerce Research*, 23(4), 2737-2757, DOI: 10.1007/s10660-022-09560-w.
- Chen, J., Ma, L., Li, X., Xu, J., Cho, J.H., Nag, K. and Achan, K. (2024). "Relation labeling in product knowledge graphs with large language models for e-commerce," *International Journal of Machine Learning and Cybernetics*, 15(12), 5725-5743, DOI: 10.1007/s13042-024-02274-5.
- Coban, O., Yağanoğlu, M. and Bozkurt, F. (2024). "Domain Effect Investigation for Bert Models Fine-Tuned on Different Text Categorization Tasks," *Arabian Journal for Science and Engineering*, 49(3), 3685-3702, DOI: 10.1007/s13369-023-08142-8.
- Gürbüz, M. and Kotan, M. (2025). "Multi-category e-commerce insights via social media analysis using machine learning and BERT," *Acta Infologica*, 9(1), 1-18, DOI: 10.26650/acin.1483488.
- El-Ansari, A. and Beni-Hssane, A. (2023). "Sentiment Analysis for Personalized Chatbots in E-Commerce Applications," *Wireless Personal Communications*, 129(3), 1623-1644, DOI: 10.1007/s11277-023-10199-5.
- Gardazi, N.M., Daud, A., Malik, M.K., Bukhari, A., Alsahfi, T. and Alshemaimri, B. (2025). "BERT applications in natural language processing: a review," *Artificial Intelligence Review*, 58(6), 1-49, DOI: 10.1007/s10462-025-11162-5.
- Goud, A. and Garg, B. (2025). "A novel framework for aspect based sentiment analysis using a hybrid BERT (HybBERT) model," *Multimedia Tools and Applications*, 84(29), 34819-34851, DOI: 10.1007/s11042-023-17647-1.
- Hao, S., Zhang, P., Liu, S. and Wang, Y. (2023). "Sentiment recognition and analysis method of official document text based on BERT-SVM model," *Neural Computing and Applications*, 35(35), 24621-24632, DOI: 10.1007/s00521-023-08226-4.
- Hicham, N. and Nasser, H. (2024). "Improving emotion classification in e-commerce customer review analysis using GPT and meta-ensemble deep learning technique for multilingual system," *Multimedia Tools and Applications*, 83(39), 87323-87367, DOI: 10.1007/s11042-024-19965-4.
- Hung, B.T. and Huy, V.Q. (2025). "MTFIC: enhanced fashion image captioning via multi-transformer architecture with contrastive and bidirectional encodings," *Visual Computer*, 41(1), 10841-10855, DOI: 10.1007/s00371-025-04072-8.
- Kansal, R. and Diwaker, C. (2025). "Hybrid recommendation approach for sentiment analysis using text-BERT and Binary Horse Herd Optimization Algorithm (BHHA) in social networks," *International Journal of Information Technology*, 17(8), 4823-4828, DOI: 10.1007/s41870-025-02679-7.
- Karabila, I., Darraz, N., EL-Ansari, A., Alami, N. and Mallahi, M.E. (2024). "BERT-enhanced sentiment analysis for personalized e-commerce recommendations," *Multimedia Tools and Applications*, 83(19), 56463-56488, DOI: 10.1007/s11042-023-17689-5.
- Karn, A.L., Karna, R.K., Kondamudi, B.R., Bagale, G., Pustokhin, D.A., Pustokhina, I.V. and Sengan, S. (2023). "RETRACTED ARTICLE: Customer centric hybrid recommendation system for E-Commerce applications by integrating hybrid sentiment analysis," *Electronic Commerce Research*, 23(1), 279-314, DOI: 10.1007/s10660-022-09630-z.
- Kassem, M.A., Abohany, A.A., El-Mageed, A.A.A. and Hosny, K.M. (2024). "A novel deep learning model for detection of inconsistency in e-commerce websites," *Neural Computing and Applications*, 36(17), 10339-10353, DOI: 10.1007/s00521-024-09590-5.
- Koy, A. and Çolak, A.B. (2024). "The intraday high-frequency trading with different data ranges: A comparative study with artificial neural network and vector autoregressive models," *Archives of Advanced Engineering Science*, 2(3), 123-133, DOI: 10.47852/bonviewAAES32021325.
- Kumar, R., Mukherjee, S. and Choudhary, D. (2025). "Uncovering the roots of customer dissatisfaction via Amazon reviews: a hybrid ensemble-deep learning approach for E-commerce quality management," *Annals of Operations Research*, 353(8), 545-574, DOI: 10.1007/s10479-025-06770-x.
- Li, X., Ma, J. and Li, S. (2025). "Intelligence customs declaration for cross-border e-commerce based on the multi-modal model and the optimal window mechanism," *Annals of Operations Research*, 348(1), 3-27, DOI: 10.1007/s10479-022-04799-w.
- Ma, J., Guo, X. and Zhao, X. (2024). "Identifying purchase intention through deep learning: analyzing the Q andD text of an E-Commerce platform," *Annals of Operations Research*, 339(1), 329-348, DOI: 10.1007/s10479-022-04834-w.
- Maiga, S., Bilgaiyan, S. and Sagnika, S. (2025). "Predicting software effort using BERT-based word embeddings," *International Journal of System Assurance Engineering and Management*, 16(5), 1728-1742, DOI: 10.1007/s13198-025-02746-z.
- Noorian, A., Harounabadi, A. and Hazratifard, M. (2024). "A sequential neural recommendation system exploiting BERT and LSTM on social media posts," *Complex & Intelligent Systems*, 10(1), 721-744, DOI: 10.1007/s40747-023-01191-4.
- Peng, Q., Luo, X., Yuan, Y., Gu, F., Shen, H. and Huang, Z. (2025). "Text classification methods in the steel e-commerce industry using the large language model and prompt engineering," *International Journal of Machine Learning and Cybernetics*, 16(11), 9399-9412, DOI: 10.1007/s13042-025-02760-4.

- Pinar, A. (2025). "An integrated sentiment analysis and q-rung orthopair fuzzy MCDM model for supplier selection in E-commerce: a comprehensive approach," *Electronic Commerce Research*, 25(2), 1311-1342, DOI: 10.1007/s10660-023-09768-4.
- Yan, J. (2024). "Data privacy regulation and cross-border e-commerce," *Empirica*, 51(4), 913-927, DOI: 10.1007/s10663-024-09624-0.
- Teke, B., Zamir, G., Budak, A. B. and Aksakallı, I. K. (2025). "BERTurk-Based Sentiment Analysis on E-Commerce Multi Domain Product Reviews," *Afyon Kocatepe Üniversitesi Fen Ve Mühendislik Bilimleri Dergisi*, 25(3), 497-509, DOI: 10.35414/akufemubid.1537513.
- Zardak, S.R., Rasekh, A.H. and Bashkari, M.S. (2023). "Persian Text Sentiment Analysis Based on BERT and Neural Networks," *Iranian Journal of Science and Technology, Transactions of Electrical Engineering*, 47(4), 1623-1634, DOI: 10.1007/s40998-023-00626-5.
- Zhao, Z., Zhang, N., Xiong, J., Feng, M., Jiang, C. and Wang, X. (2024). "Enhancing E-commerce Recommendations: Unveiling Insights from Customer Reviews with BERTFusionDNN," *Journal of Theoretical and Practical Engineering Science*, 4(2), 38-44, DOI: 10.53469/jtpes.2024.04(02).06.
- Zou, W., Zhang, W., Wu, W. and Tian, Z. (2024). "A Multi-task Shared Cascade Learning for Aspect Sentiment Triplet Extraction Using BERT-MRC," *Cognitive Computation*, 16(4), 1554-1571, DOI: 10.1007/s12559-024-10247-7



Danyan Chen obtained her master's degree in English language and literature (2006) from NNU, China. Presently, she is working as an Associate Professor and Dean in the teaching and research office for Cross-border Business at Wuxi Vocational Institute of Commerce, China. She was appointed to the Talent Development Special Committee of the National Cross-border E-commerce Industry-Education Integration Consortium. She has published more than 20 articles in reputed journals. Her areas of interest include cross-border e-commerce, business English and higher vocational education.