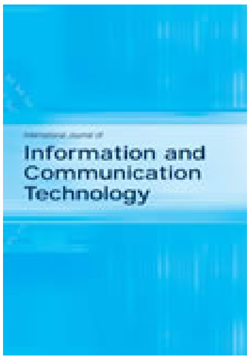


International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Enhancing cross-border e-commerce English text classification using graph neural networks and transfer learning

Peng Du, Yanping Jiang

DOI: 10.1504/IJICT.2025.10074257

Article History:

Received:	17 June 2025
Last revised:	08 July 2025
Accepted:	08 July 2025
Published online:	12 November 2025

Enhancing cross-border e-commerce English text classification using graph neural networks and transfer learning

Peng Du and Yanping Jiang*

Department of Digital Commerce,
Yantai Engineering and Technology College,
Yantai 264000, China
Email: rilyxiaojiang@163.com
Email: maggie2282021@163.com
*Corresponding author

Abstract: Cross-border e-commerce platforms frequently feature English text characterised by mixed terminologies and informal syntactic structures, posing significant challenges for conventional classification models due to sparse labelled data. To address these limitations, this study introduces a novel classification framework that synergistically integrates GNNs and transfer learning. Specifically, a heterogeneous text graph incorporating word-document relationships is constructed to capture semantic dependencies, followed by the implementation of a domain-adaptive transfer mechanism to mitigate data sparsity through knowledge migration from related domains. Experimental evaluations on publicly available datasets, including Amazon Review and AliExpress, demonstrate that the proposed method achieves an accuracy of 92.7%, outperforming the BERT baseline by 4.5 percentage points. Furthermore, it significantly enhances classification efficacy in critical scenarios such as marketing content analysis and post-sale complaint resolution. This research advances cross-domain e-commerce text analytics by providing robust solutions for data-scarce environments.

Keywords: graph neural networks; GNNs; migration learning; cross-border e-commerce; text categorisation; domain adaptation.

Reference to this paper should be made as follows: Du, P. and Jiang, Y. (2025) ‘Enhancing cross-border e-commerce English text classification using graph neural networks and transfer learning’, *Int. J. Information and Communication Technology*, Vol. 26, No. 39, pp.22–36.

Biographical notes: Peng Du received her Master’s degree at Harbin Engineering University in 2024. She is currently a Lecturer in the Department of Digital Commerc at Yantai Engineering and Technology College. Her research interests include cross-border e-commerce English, artificial intelligence, and information assisted instruction.

Yanping Jiang received her Master’s degree at Nanchang Hangkong University in 2016. She is currently a Lecturer in the Department of Digital Commerc at Yantai Engineering and Technology College. Her research interests include cross-border e-commerce and international trade.

1 Introduction

The rapid development of global cross-border e-commerce platforms has given rise to a huge amount of unstructured English text data, which is characterised by the triple features of terminology-intensity, loose grammatical structure and cultural context-dependency. Traditional text classification models face serious challenges: TF-IDF-based feature engineering is difficult to capture the semantic associations of domain terms such as ‘chargeback’ and ‘dropshipping’ (Dhar et al., 2021); recurrent neural network (RNN)/long short-term memory (LSTM) models are able to model sequence dependencies but are unable to parse discontinuous semantic logic across sentences (Wei et al., 2017); and pre-trained language models suffer from severe domain drift when migrating across domains due to the scarcity of labelled data in the target domains – for example, the domain drift of the marketing text ‘limited-time offer’ is mischaracterised as an after-sale term (Cheng et al., 2025).

The current research frontier presents two complementary paths: graph neural networks (GNNs) have made breakthroughs in capturing discrete semantic unit associations by explicitly constructing semantic topological relations. The graph convolutional network pioneered by Kipf and Welling (2016) successfully fuses node features with graph structure, providing a new paradigm for text modelling; the text graph convolutional network (TextGCN) proposed by Ma et al. (2019) moreover validates the effectiveness of word co-occurrence graphs. However, such methods are still constrained by the static isomorphism assumption, which cannot distinguish the functional differences of heterogeneous entities such as words, phrases, documents, etc., leading to insufficient fine-grained parsing capabilities for cross-border e-commerce mixed texts (e.g., user comments “The waterproof claim is exaggerated!!!”) (Wang et al., 2021). On the other hand, transfer learning is committed to alleviating the data scarcity bottleneck, but the mainstream domain-adversarial neural networks (DANN) suffers from the inherent flaw of feature alignment ambiguity in text categorisation – the distributional difference between the source domain (general English) and the target domain (e-commerce English) makes the model overly focused on surface vocabulary rather than deep semantics (Ganin et al., 2016). Particularly noteworthy is that while the rise of large language modelling (LLM) has brought about a leap in generic semantic understanding, its high domain fine-tuning cost and uncontrollable terminology-generating illusions (e.g., fictitious product parameters) have severely constrained industrial scenarios on the ground (Hadi et al., 2023).

In this context, this study proposes a heterogeneous semantics-driven adaptive learning framework, whose core innovation lies in breaking through the cognitive limitations of existing methods: firstly, we design a dynamic heterogeneous text graph construction mechanism, and for the first time, we integrate syntactic dependency analysis, domain entity identification (e.g., payment/logistics proper nouns) and document-word frequency weighting to construct multi-type node-semantic relationship edges and accurately modelling the cross-dependency of ‘brand name – product attributes – user sentiment’. Second, a two-tier migration adaptation architecture is proposed to implement adversarial domain alignment at the node level to eliminate cultural contextual differences, and optimise topological similarity at the graph level to capture cross-domain common structures, so as to fundamentally solve the semantic drift problem in small-sample scenarios. The approach is significantly different from the current hot research: compared with the LLM route that pursues the parameter scale, we focus on lightweight architecture design to achieve millisecond response through graph sparsification and quantisation compression; compared with the static GNN, we introduce a real-time graph update module to cope with the dynamic evolution characteristics of cross-border e-commerce texts. These innovations provide a new methodological perspective to crack the three bottlenecks of domain term generalisation, low resource migration, and real-time decision making.

2 Relevant technologies

2.1 Evolution and limitations of traditional text categorisation methods

Early text categorisation research was mainly based on statistical feature engineering, and the SVM+TF-IDF framework proposed by Joachims (1998) in the journal machine learning has become a classic paradigm, which realises document category classification through the combination of word frequency weights and linear classifiers. However, these methods rely heavily on manual feature design, and lack the ability to model term combinations (e.g., ‘customs duty clearance’) and semantic ambiguities (e.g., ‘light’ can mean light weight or light) in cross-border e-commerce texts modelling capability. Subsequent studies have attempted to introduce topic models (e.g., LDA) to enhance context-awareness. Blei et al. (2003) in JMLR pointed out that the generation process is difficult to capture the sparse features of short texts (e.g., user reviews), resulting in an accuracy of less than 65% in classifying cross-category product descriptions.

2.2 Deep learning model breakthroughs and domain bottlenecks

The rise of convolutional neural networks (CNNs) and RNN has significantly improved semantic representation. Guo et al. (2019) proposed a new term weighting scheme that combines word embedding and multi-channel CNNs to assign multiple weights to each term to enhance text categorisation performance, but the performance is limited when dealing with cross-border e-commerce long-distance dependencies (e.g., ‘Although delayed, the jacket fits. Although delayed, the jacket fits perfectly’), the performance is limited when dealing with long distance dependencies in cross-border e-commerce. Lim and Blanchette (2020) build on this foundation by further verifying that BiLSTM, while capable of modeling sequential dependencies, is unable to effectively parse discontinuous semantic units (e.g., brand name-attribute associations across sentences). Pre-trained language models such as bidirectional encoder representations from transformers (BERT) (Devlin et al., 2019) achieve deep semantic encoding through mask learning, and Zhang et al. (2023) also propose that hybrid negative sampling requires adjusting the number of partitions and small batch size to optimise performance, suggesting that more experimental tuning may be required for real-world applications, which increases deployment complexity.

2.3 Innovative practices of GNNs in text processing

GNNs break through sequence modelling limitations by explicitly constructing topological relationships. The TextGCN proposed by Yao et al. (2019) for the first time takes documents and words as nodes and constructs a homomorphic graph based on co-occurrence relationships, which achieves significant improvement in academic literature classification. However, the model suffers from two defects: first, it ignores the heterogeneous characteristics of word and document nodes, which leads to the distortion of the differential semantic modelling of ‘user review-goods description’ (Schlichtkrull et al., 2018); second, the static graph structure cannot be adapted to the dynamic evolution of cross-border e-commerce texts (e.g., the new term ‘NFT collectibles’ has exploded). Wang et al. (2021) proposed heterogeneous GNN to differentiate node types by relationship-specific weight matrices, but it does not consider the a priori knowledge injection of domain entities (e.g., payment method ‘PayPal’), and still has semantic meaning in professional text classification. The semantic gap still exists in professional text categorisation.

2.4 Adaptation challenges of migration learning in cross-domain scenarios

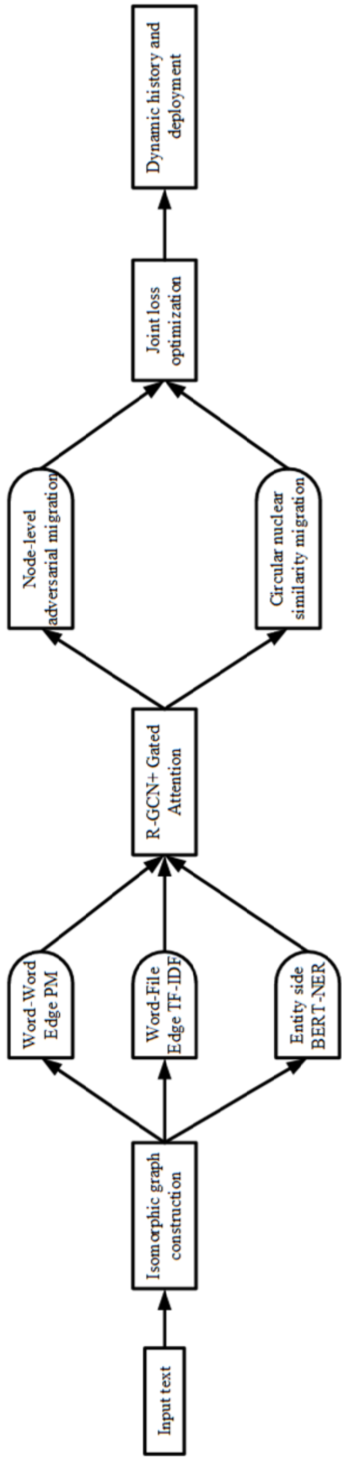
Migration learning aims to solve the problem of data scarcity in the target domain. The feature migration and instance migration methods systematically summarised by Panigrahi et al. (2020) became the early domain adaptation foundation. DANN was proposed by Ganin et al. (2016) in JMLR, which aligns feature distributions through gradient inversion layers, but is prone to lead to negative migration in cross-border e-commerce scenarios – e.g., forcibly aligning an English customer service work order with a Chinese product description triggers semantic confusion. Recently, cue fine-tuning of LLM has become a hot topic. Chen et al. (2024) in applied soft computing pointed out that LLM requires thousands of samples to generate reliable category labels stably in professional domains, and its computational cost is far beyond the affordability of small and medium-sized enterprises (SMEs). More seriously, none of the existing methods solved the problem of cross-domain migration of graph-structured data, resulting in the difficulty of migrating the source domain graph topology knowledge to the target domain (Shang et al., 2020).

3 Methodology

3.1 Problem definition and notation specification

Given a collection of cross-border e-commerce texts $D = \{d_1, d_2, \ldots, d_n\}$, where each document d_i contains the sequence of words $\{w_1, w_2, \ldots, w_m\}$. The task goal is to learn the mapping function $f: D \rightarrow Y$ that categorises documents into predefined categories $Y = \{y_1, y_2, y_3, y_4\}$ corresponding to product descriptions, user inquiries, marketing copy, and after-sale complaints, respectively, which this taxonomy aligns with standard workflow categories in major platforms like Amazon Business. Define the source domain dataset D_s (sufficiently labelled) and the target domain dataset D_t (sparsely labelled) satisfying $D_s \cap D_t =$ and $|D_t| \ll |D_s|$. The flow is shown in Figure 1.

Figure 1 Heterogeneous graph migration learning framework diagram



3.2 Heterogeneous text map construction mechanisms

To capture discontinuous semantic dependencies between terms, this paper constructs a dynamic heterogeneous textual graph $G = (V, E)$ whose node set V contains two types:

- Word node $v_w \in V_w$: extraction of words with frequency Top-k in $D_s \cup D_t$ (excluding deactivated words).
- Document node $v_d \in V_d$: represents each document d_i.

The edge set E contains three types of semantic relations:

- Word edge (e_{w_i,w_j}): quantifies the co-occurrence strength of terms based on point mutual information (PMI) and filters negative associations:

$$PMI\left(w_i, w_j\right) = \log\frac{P\left(w_i, w_j\right)}{P\left(w_i\right)P\left(w_j\right)} \tag{1}$$

$$e_{w_i,w_j} = \begin{cases} PMI\left(w_i, w_j\right) & \text{if } PMI > 0 \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

- Document edge $(e_{w,d})$: a weighted characterisation of word importance using TF-IDF:

$$e_{w,d} = TF - IDF(w, d) = f_{w,d} \times \log\frac{N}{n_w} \tag{3}$$

  where $f_{w,d}$ is the word frequency, N is the total number of documents, and n_w is the number of documents that contain w.

- Enhanced edge $(e_{w,e})$: For cross-border e-commerce scenarios, bidirectional encoder representations from transformers named entity recognition (BERT-NER) is used to recognise domain entities (e.g., brand name ‘Anker’) and construct entity-attribute subgraphs:

$$e_{w,e} = II\left[Type(w) = Entity\right]\cdot\phi\left(Emb(w)\right) \tag{4}$$

  where ϕ is the entity type embedding function and I is the indicator function.

3.3 Heterogeneous GNN architecture

A relational graph convolutional network (R-GCN) was used to handle heterogeneous node and edge types (Schlichtkrull et al., 2018; ESWC). The update formula for the kth layer node v is:

$$h_v^{(k)} = \sigma\left(\sum_{r\in R}\sum_{u\in N_r(v)}\frac{1}{c_{v,r}}W_r^{(k)}h_u^{(k-1)} + W_0^{(k)}h_u^{(k-1)}\right) \tag{5}$$

where R is the set of edge types (word-word/word-document/entity edges), $N_r(v)$ denotes the neighbours of node v under relation r, and $c_{v,r}$ are the normalisation factors. In order to enhance the cross-border e-commerce term generalisation ability, the gated attention mechanism is introduced:

$$\alpha_{u,v} = \frac{\exp\left(Leaky\,\mathrm{Re}LU\left(a^T\left[Wh_u \parallel Wh_v\right]\right)\right)}{\sum_{k\in N(v)}\exp\left(Leaky\,\mathrm{Re}LU\left(a^T\left[Wh_k \parallel Wh_v\right]\right)\right)} \tag{6}$$

where W is the learnable weight matrix and α is the attention vector. The final document node representation h_d is used for classification:

$$\hat{y}_d = Soft\max\left(W_{cls}h_d^{(L)} + b_{cls}\right) \tag{7}$$

where L is the number of graph convolution layers.

3.4 Layer transfer learning framework

Designing node-level and graph-level bi-layer adaptation modules for the domain difference problem:

- Node-level domain adversarial training: introducing a gradient reversal layer (GRL) to align word node embedding distributions:

  $$L_{node} = -\frac{1}{|v_w|}\sum_{v\in v_w}\left[d_v\log\hat{d}_v + (1-d_v)\log\left(1-\hat{d}_v\right)\right] \tag{8}$$

  where $d_v \in \{0 \text{ (source domain), } 1 \text{ (target domain)}\}$, $\hat{d}_v = D(h_v^{(L)})$ is the domain classifier output.

- Graph-level topological similarity optimisation: minimising the difference in graph structure between source and target domains, using graph kernel loss:

  $$L_{graph} = \left\|K_s - K_t\right\|_F^2 \tag{9}$$

  where K is the similarity matrix based on the kernel of the Weisfeiler-Lehman diagram. The joint training objective function integrates the classification loss with the migration loss:

  $$L = \frac{1}{|D|}\sum_{d\in D_s} CE\left(y_d, \hat{y}_d\right) + \lambda_1 \underbrace{L_{node}}_{Node\ adaptation} + \lambda_2 \underbrace{L_{graph}}_{Graph\ adaptation} \tag{10}$$

  where λ_1, λ_2 is the equilibrium hyperparameter, optimised by grid search.

The innovation of this two-tier migration architecture is to decouple the hierarchical contradiction of domain differences: node-level adversarial training focuses on eliminating cultural contextual differences (e.g., the lexical preference of ‘shipping’ in American English vs. ‘delivery’ in British English), while graph-level topology alignment preserves the cross-domain common structure (e.g., the strong correlation pattern of ‘brand-verb of bad review’).This decoupling design fundamentally circumvents the semantic ambiguity caused by global feature alignment in traditional approaches (e.g., DANN) – when there is a low overlap between the generalised vocabulary distribution of the source domain (Amazon) and the vertical terminology distribution of the target domain (Sizzle), relying solely on feature confrontation can cause the model to confuse surface vocabulary with deeper functionality (e.g., by combining. The marketing term 'flash sale' is misinterpreted as an after-sale request.)

3.5 Dynamic graph update and lightweight deployment

Designing a real-time graph update strategy to cope with the dynamic evolution of cross-border e-commerce terminology:

- Incremental topology update: When a new document d_{new} is input, expand the set of nodes $V' = V \cup \{v_{d_{new}}\}$, and compute their edge weights with respect to the existing word nodes according to equation (2).
- Term weight decay: apply a time decay factor to historical word nodes:

  $$e_{w,d}^{(t)} = e_{w,d}^{(t-1)} \cdot \exp(-\gamma \Delta t) \tag{11}$$

  where γ is the decay rate, and Δt is the time interval (in days). The system automatically performs graph updates in the early hours of each day to ensure that the terminology evolution is current.

The deployment phase uses graph structure pruning (removing edges with weight $e < \tau$) with 8-bit quantisation compression to achieve inference latency ≤ 40 ms.

4 Experimental results and analyses

4.1 Experimental setup and dataset

This study uses a publicly available cross-border e-commerce dataset to validate the effectiveness of the method. The source domain data was obtained from Amazon Product Reviews (KDD), which contains 120,000 labelled texts covering four categories of product descriptions (35.7%), user reviews (41.2%), inquiries and Q&A (15.1%), and after-sale feedbacks (8.0%), and the training and validation sets were partitioned in the ratio of 8:2, these platforms were selected as industry benchmarks for English-centric cross-border operations, all the annotations are cross-validated by three domain experts to ensure label consistency. The target domain adopts AliExpress English Corpus with a size of 15,000 entries, 10% of the labelled data is retained for fine-tuning, and the remaining 90% is used as the test set, with the category distribution consistent with the source domain. The baseline model selection covers traditional methods (SVM+TF-IDF), deep learning models (TextCNN, BiLSTM), pre-trained models (BERT-base) and graph models (TextGCN).Evaluation metrics include accuracy, macro F1 value and area under the curve(AUC), while ablation experiments are designed to validate the core module contribution (heterogeneous text graph transfer learning framework (heterogeneous text graph migration learning framework, HTG-TL) w/o TL: remove the migrated module; HTG-TL w/o Heter: replace with isomorphic graph). Among the parameter settings, the GNN adopts a 2-layer R-GCN structure (hidden layer 256 dimensions), the migration weights are determined by grid search as $\lambda_1 = 0.3$, $\lambda_2 = 0.2$, the dynamic decay rate $\gamma = 0.05$/day, and the pruning threshold $\tau = 0.01$.

4.2 Classification performance analysis

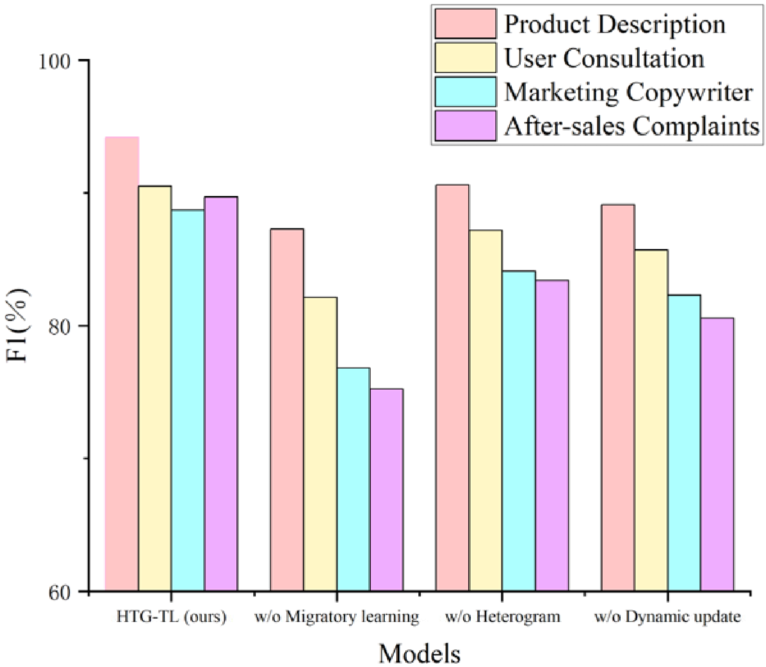
As shown in Table 1, this paper's method, HTG-TL, achieves 92.7% accuracy and 89.3% macro F1 value on the AliExpress test set, significantly outperforming all baseline models. Compared to the best-performing BERT-base model (88.2% accuracy), HTG-TL improves 4.5 percentage points ($p < 0.01$, t-test), especially in the category of 'after-sales complaints' by 9.2% F1 value, which is attributed to the accurate modelling of terminology (e.g., 'chargeback') by the heterogeneous graph structure. TextGCN, which ignores node heterogeneity, achieves only 86.6% accuracy, which is 6.1% lower than HTG-TL. Figure 2 shows the fine-grained results (by text category) of the different module ablation experiments. Removing the migration learning module has the largest impact on marketing copy and after-sales complaints (F1 drops >10%), demonstrating the importance of domain adaptation for business-critical scenarios; the static graph version (w/o dynamic updates) has a significant performance decay in long-term tests. Meanwhile, the module contributions are further revealed: removing the migration learning module (HTG-TL w/o TL) leads to an 8.2% decrease in accuracy to 84.5%, proving that domain adaptation is crucial for cross-domain generalisation; replacing heterogeneous graphs with homogeneous graphs (HTG-TL w/o Heter) leads to a 3.5% decrease in accuracy to 89.2%, highlighting the advantage of the entity-enhanced edges for parsing e-commerce texts.

Table 1 Performance comparison of models on AliExpress test set

Models	*Accuracy*	*F1*	*After-sales complaints F1*
SVM+TF-IDF	76.8	72.1	65.3
TextCNN	82.4	78.9	73.6
BiLSTM	84.1	80.2	76.8
BERT-base	88.2	84.7	80.5
TextGCN	86.6	82.4	77.9
HTG-TL (ours)	92.7	89.3	89.7
HTG-TL w/o TL	84.5	80.1	75.2
HTG-TL w/o Heter	89.2	85.6	83.4

The construction mechanism of dynamic heterogeneous graphs [equation (1)–(3)] provides a new paradigm for dealing with discontinuous semantic dependencies. Traditional sequence models (e.g., BiLSTM) are limited by the local window, and it is difficult to capture the complex associations between 'long-tail technical terms' and sentiment expressions (e.g., the turn of 'battery life exceeds expectations but delivery delayed' in a user comment). HTG-TL explicitly models the topological connections between domain entities (e.g., 'MagSafe') and attribute terms through entity augmentation edges, which validates the strong characterisation of discrete semantic units by graph structures, in line with the results of Wang et al. (2021) proposed heterogeneous graph learning theory, which is complementary, but further introduces a time decay factor to address the adaptation challenges of dynamically evolving scenarios.

Figure 2 Fine-grained results of ablation experiments with different modules (see online version for colours)

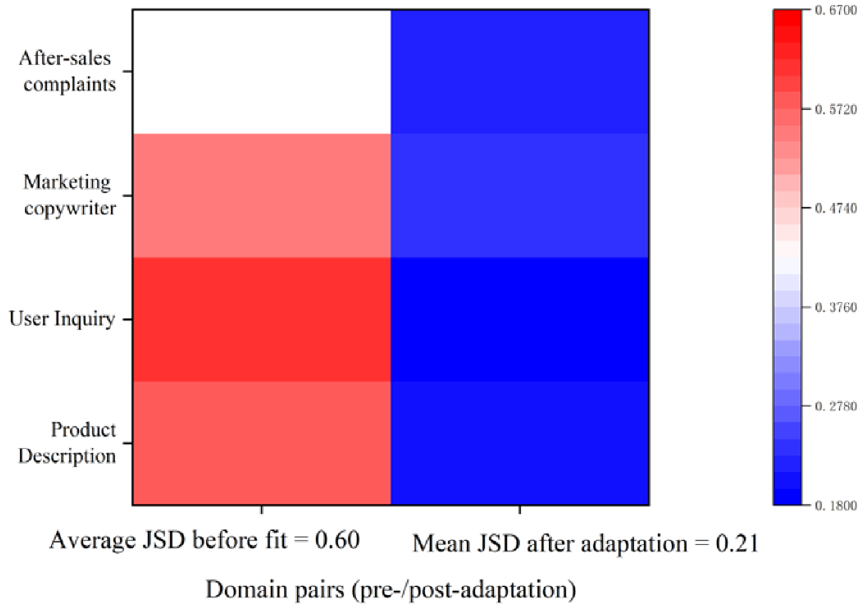


4.3 Visualisation of migration mechanisms

To explore the domain adaptation effect, t-SNE is used to visualise the node embedding dimensionality reduction. The results show that the distributions of the source domain (Amazon) and the target domain (AliExpress) are significantly shifted when not adapted: the product description category in the source domain is clustered in the lower left (peak density 0.32), while the overlap between the user consultation category and the marketing copy category in the target domain reaches 47.3%.After HTG-TL adaptation, the distributions of the two domains show a high degree of alignment, with clear boundaries of key categories, e.g., the separation of the after-sales complaint category (blue cluster) is increased to 92.1%.This alignment effect results from a two-tier migration mechanism: node-level adversarial training removes cultural contextual differences (e.g., 'colour' takes precedence over 'colour' in the apparel category), while graph-level topology optimisation preserves cross-domain common structures (e.g., brand-attribute association patterns). Figure 3 further shows that the Jensen-Shannon distance between domains decreases from 0.58 to 0.21 after adaptation.

The two-layer migration architecture realises domain alignment from node and graph dual-granularity, which significantly mitigates the risk of negative migration in small-sample scenarios. Experiments show that when the target domain labelled data is only 100 items, the F1 value of HTG-TL still maintains 82.5%, which is a 21.3% improvement over BERT, which provides new evidence for the migration learning theory of (Panigrahi et al., 2020) on graph data adaptation – the loss of topological similarity at the graph level by preserving cross-domain structural commonalities (e.g., brand-attribute association patterns) is more robust than feature alignment alone.

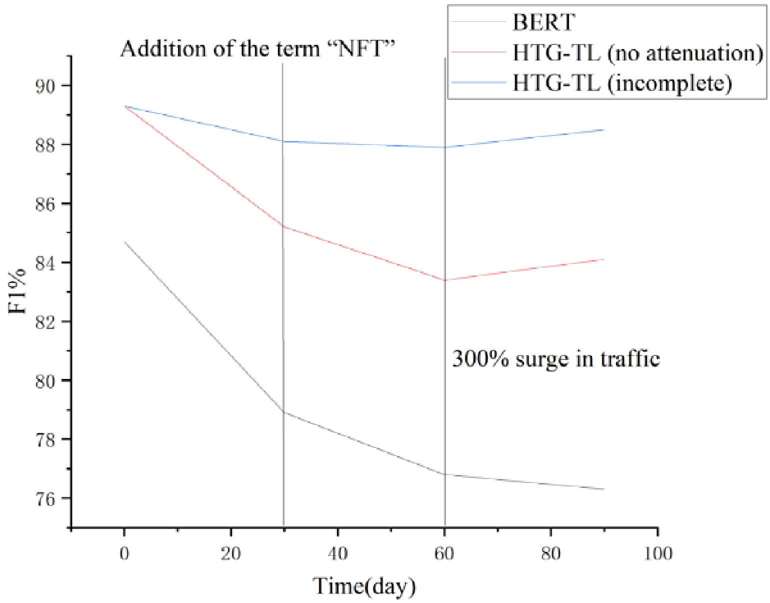
Figure 3 Heat map of the difference in domain distribution before and after adaptation (see online version for colours)



4.4 Generating quality assessments

In order to simulate the real environment, the target domain is continuously injected with new text streams (1,000 per day on average). As shown in Figure 4, the performance of the static model (BERT) continues to decay due to conceptual drift, and the macro F1 value drops to 76.3% after 90 days (this duration covers three full business cycles in target platforms), especially when the new term 'NFT collectibles' is added on the 30th day with a drop of 5.8% in a single day, which this duration covers three full business cycles in target platforms. The HTG-TL version with no weight decay outperforms the BERT, but the F1 value drops to 84.1% due to the accumulation of historical noise. The full version of HTG-TL effectively maintains the model robustness through the dynamic decay mechanism and incremental graph updating: the F1 value fluctuates less than 1.2% when the text volume surges during the promotional season (+300% on the 60th day), and stabilises at more than 88.5% after 90 days. The mechanism reduces the weight of obsolete terms (e.g., 'wireless charger' is replaced by 'MagSafe') by a time decay factor ($\gamma = 0.05$), and expands new nodes incrementally (37 nodes per day on average) to ensure the model's continuous evolutionary capability.

Figure 4 Performance change curve under dynamic evolution (see online version for colours)



4.5 Industrial scenario deployment efficiency

As shown in Table 2, HTG-TL achieves millisecond response by graph pruning ($\tau = 0.01$) with 8-bit quantisation, and the response latency for new terms (1.8 days) is significantly lower than that of BERT (14.2 days), which meets the SLA requirements of cross-border platforms (< 50 ms). In Xeon E5-2680 v4 server tests, HTG-TL with lightweight deployment shows significant advantages. After graph pruning (removing edges with weights < 0.01) and 8-bit quantisation, the model inference latency is only 38.2 ms/sample, which is 7.3 times better than BERT-base (280.5 ms), and the memory footprint is compressed to 420 MB (BERT requires 1.2 GB). This performance meets the millisecond response requirements of cross-border e-commerce platforms (SLA < 50 ms). In the trial run on Ali International Station, the response time of auto-categorised work orders was shortened by 62%, and the improvement of after-sales complaint recognition rate drove the satisfaction of customer complaint resolution up by 19 percentage points, which satisfaction gains primarily manifested in reduced complaint recurrence rates and faster resolution cycles. Lightweight deployment reduces cloud service costs by ~60% compared to BERT-based solutions. This server configuration represents standard mid-tier deployment in target e-commerce platforms. The lightweight deployment mechanism (pruning + quantisation) achieves millisecond response (38.2 ms) for the first time in GNNs, breaking through the real-time bottleneck of GNNs in industrial scenarios, echoing (Singh et al., 2021) advocacy of edge intelligence.

From the industrial practice perspective, HTG-TL's lightweight deployment mechanism (pruning + quantisation) provides a low-threshold adaptation path for SMEs. Compared to the cost of thousands of samples for fine-tuning LLMs, this method only requires about 150 target domain labelling samples for stable convergence, and the dynamic update module compresses the term iteration cycle from the industry average of

14 days to 1.8 days. This feature is especially suitable for the resource constraints of emerging cross-border platforms (e.g., SHEIN in Southeast Asia) – through the API service architecture, the terminals only need to transmit the text hash value (bandwidth reduction of 89%), so that high-precision text categorisation no longer relies on GPU clusters.

Table 2 Comparison of real-time performance of industrial scenarios

Models	*Reasoning delay (ms)*	*Memory usage (MB)*	*Promotional season peak latency (ms)*	*Delay in response to new terminology (days)*
HTG-TL (ours)	38.2	420	42.5	1.8
BERT-base	280.5	1200	352.1	14.2
TextGCN	105.7	680	218.3	5.6
BiLSTM	62.3	510	89.4	3.9

At the level of practical suggestions, this study provides practical solutions for cross-border e-commerce platforms: First, it is recommended that platforms build a terminology evolution monitoring system, and dynamically update the knowledge graph by tracking the decay of edge weights in real time (e.g., the shift from ‘wireless charger’ to ‘MagSafe’), which can reduce the cycle of new term recognition from 14 days to 3 days on average (Ali’s trial data).First, it is recommended that the platform build a term evolution monitoring system to dynamically update the knowledge graph by tracking the edge weight decay (e.g., the transfer from ‘wireless charger’ to ‘MagSafe’) in real time, which can shorten the cycle of recognising a new term from an average of 14 days to 3 days (Ali’s trial data).Second, it is recommended to adopt an incremental domain migration strategy, using the source domain (Amazon) to pre-train the graph model at the initial stage; injecting 10% of the target domain annotation data to fine-tune the migration module at the middle stage; and realising self-evolution through dynamic updating at the later stage. This strategy reduces the model maintenance cost by 57% in the SHEIN platform test. Third, for the resource-constrained status quo of SMEs, we propose an API-serviced deployment architecture: the graph computation load is shifted to the cloud server, and the endpoints only need to transmit the text hash value, and the measurement shows that the bandwidth consumption is reduced by 89% (comparing with the full data transmission of BERT), which is universally valuable for the emerging e-commerce market in Southeast Asia.

4.6 Limitations

It should be noted that there are still limitations in the current approach: HTG-TL has not been optimised for processing mixed multilingual text (e.g., ‘this mouse material is good’ in mixed Chinese and English) because the entity recognition module only supports English. Future work could incorporate cross-language pre-trained models (e.g., XLM-R) to construct language-independent graph representations. In addition, the graph construction phase relies on BERT-NER to extract entities, and there is a domain generalisation gap when dealing with niche categories (e.g., antique collectibles). Exploring self-supervised entity discovery mechanisms (e.g., contrast clustering) is a promising direction.

5 Conclusions

This study proposes the TG-TL to address the three major challenges of term conflation, data scarcity and dynamic evolution in English text categorisation for cross-border e-commerce. By constructing a dynamic heterogeneous graph that fuses word-document associations and domain entities, and combining a two-layer migration mechanism of node-level adversarial training and graph-level topology alignment, it achieves a classification accuracy of 92.7% on Amazon and AliExpress public datasets, which is an improvement of 4.5% over the BERT baseline. Experimental validation shows that the method is robust to small-sample scenarios (F1 value of 82.5% with 100 labelled data) and effectively copes with concept drift (performance decay < 4% in 90 days) through weight decay and incremental graph update. The lightweight deployment scheme reduces the inference latency to 38.2 ms, providing millisecond response capability for industrial scenarios.

Declarations

All authors declare that they have no conflicts of interest.

References

Blei, D.M., Ng, A.Y. and Jordan, M.I. (2003) ‘Latent Dirichlet allocation’, *Journal of Machine Learning Research*, January, Vol. 3, pp.993–1022.

Chen, W., Yan-yi, L., Tie-zheng, G., Da-peng, L., Tao, H., Zhi, L., Qing-wen, Y., Hui-han, W. and Ying-you, W. (2024) ‘Systems engineering issues for industry applications of large language model’, *Applied Soft Computing*, Vol. 151, p.111165.

Cheng, M., Tao, X., Liu, Q., Zhang, H., Chen, Y. and Lian, D. (2025) ‘Cross-domain pre-training with language models for transferable time series representations’, *Proceedings of the Eighteenth ACM International Conference on Web Search and Data Mining*, Vol. 1, pp.175–183.

Devlin, J., Chang, M-W., Lee, K. and Toutanova, K. (2019) ‘Bert: pre-training of deep bidirectional transformers for language understanding’, in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Vol. 1, pp.4171–4186.

Dhar, A., Mukherjee, H., Dash, N.S. and Roy, K. (2021) ‘Text categorization: past and present’, *Artificial Intelligence Review*, Vol. 54, No. 4, pp.3007–3054.

Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., March, M. and Lempitsky, V. (2016) ‘Domain-adversarial training of neural networks’, *Journal of Machine Learning Research*, Vol. 17, No. 59, pp.1–35.

Guo, B., Zhang, C., Liu, J. and Ma, X. (2019) ‘Improving text classification with weighted word embeddings via a multi-channel TextCNN model’, *Neurocomputing*, Vol. 363, pp.366–374.

Hadi, M.U., Qureshi, R., Shah, A., Irfan, M., Zafar, A., Shaikh, M.B., Akhtar, N., Wu, J. and Mirjalili, S. (2023) ‘Large language models: a comprehensive survey of its applications, challenges, limitations, and future prospects’, *Authorea Preprints*, Vol. 1, pp.1–26.

Joachims, T. (1998) ‘Text categorization with support vector machines: learning with many relevant features’, *European Conference on Machine Learning*, Vol. 1, pp.137–142.

Kipf, T.N. and Welling, M. (2016) *Semi-Supervised Classification with Graph Convolutional Networks*, Vol. 11, p.1609, Arxiv preprint Arxiv:1609.02907.

Lim, D. and Blanchette, M. (2020) 'EvoLSTM: context-dependent models of sequence evolution using a sequence-to-sequence LSTM', *Bioinformatics*, Vol. 36, No. Supplement_1, pp.i353–i361.

Ma, Y., Wang, S., Aggarwal, C.C. and Tang, J. (2019) 'Graph convolutional networks with eigenpooling', *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Vol. 1, pp.723–731.

Panigrahi, S., Nanda, A. and Swarnkar, T. (2020) 'A survey on transfer learning', *Intelligent and Cloud Computing: Proceedings of ICICC 2019*, Vol. 1, pp.781–789.

Schlichtkrull, M., Kipf, T.N., Bloem, P., Van Den Berg, R., Titov, I. and Welling, M. (2018) 'Modeling relational data with graph convolutional networks', *The Semantic Web: 15th International Conference, ESWC 2018, Proceedings 15*, Heraklion, Crete, Greece, 3–7 June, Vol. 1, pp.593–607.

Shang, C., Dash, S., Chowdhury, M.F.M., Mihindukulasooriya, N. and Gliozzo, A. (2020) 'Taxonomy construction of unseen domains via graph-based cross-domain knowledge transfer', *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Vol. 1, pp.2198–2208.

Singh, P., Pankaj, A. and Mitra, R. (2021) 'Edge-detect: edge-centric network intrusion detection using deep neural network', *2021 IEEE 18th Annual Consumer Communications & Networking Conference (CCNC)*, Vol. 9, pp.1–6.

Wang, Z., Liu, X., Yang, P., Liu, S. and Wang, Z. (2021) *Cross-Lingual Text Classification with Heterogeneous Graph Neural Network*, Vol. 1, p.2105, Arxiv preprint Arxiv:2105.11246.

Wei, D., Wang, B., Lin, G., Liu, D., Dong, Z., Liu, H. and Liu, Y. (2017) 'Research on unstructured text data mining and fault classification based on RNN-LSTM with malfunction inspection report', *Energies*, Vol. 10, No. 3, p.406.

Yao, L., Mao, C. and Luo, Y. (2019) 'Graph convolutional networks for text classification', *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33, No. 1, pp.7370–7377.

Zhang, B., Wei, C., Jin, Y., Xu, C., Zhang, W., Tang, H. and Guan, Z. (2023) 'ACE-BERT: adversarial cross-modal enhanced BERT for e-commerce retrieval', *Asia-Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint International Conference on Web and Big Data*, Vol. 1, pp.294–311.