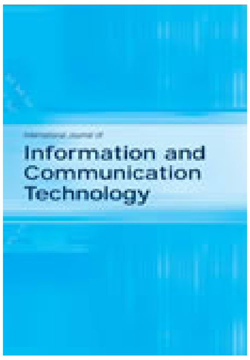


International Journal of Information and Communication Technology

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

English text classification model based on graph neural networks and contrastive learning

Yingying Cao

DOI: 10.1504/IJICT.2025.10071984

Article History:

Received:	16 May 2025
Last revised:	29 May 2025
Accepted:	29 May 2025
Published online:	10 July 2025

English text classification model based on graph neural networks and contrastive learning

Yingying Cao

Department of Education and Art,
Minxi Vocational & Technical College,
Longyan 364000, China
Email: 18760162383@163.com

Abstract: To address traditional graph neural networks' (GNNs) neglect of word-order features and sensitivity to adversarial noise, this study proposes DGCL-TC, a text classification model integrating dual-graph fusion and adaptive contrastive learning. The framework leverages bidirectional encoder representations from transformers (BERT) to encode contextual semantics and constructs dual graphs capturing local and global text structures. A learnable augmentation module dynamically generates contrastive views via node dropout and attribute masking, optimising representations through cross-view consistency. A gated graph attention network fuses topology-aware graph features with BERT embeddings, balancing structural and sequential cues. Evaluations on benchmark datasets confirm that DGCL-TC significantly outperforms baseline methods in accuracy and robustness, particularly under adversarial perturbations and sparse data conditions. The model advances text classification by unifying semantic, structural, and noise-resistant representation learning.

Keywords: text classification; graph neural network; GNNs; contrastive learning; BERT.

Reference to this paper should be made as follows: Cao, Y. (2025) 'English text classification model based on graph neural networks and contrastive learning', *Int. J. Information and Communication Technology*, Vol. 26, No. 25, pp.48–69.

Biographical notes: Yingying Cao received her Master's degree at Guangdong University of Foreign Studies in 2007. She is currently a Lecturer in the Department of Education and Art at Minxi Vocational & Technical College. Her research interests include graph neural network, computer-aided translation and large language model.

1 Introduction

With the explosive growth of information on the internet, text categorisation, as one of the core tasks of Natural Language Processing (NLP), shows irreplaceable value in the fields of opinion analysis, content recommendation, and information retrieval (Fazaeli and Momtazi, 2024). Traditional text categorisation methods rely on Bag-of-Words (BoW) or shallow neural networks such as convolutional neural network (CNN) and recurrent neural network (RNN) to extract local or sequential features are often limited by their ability to model the global semantics and structural relationships of text (Lan et al., 2023). In recent years, the rise of graph neural networks (GNNs) has provided a new perspective for text classification – by transforming text into graph structures such as word co-occurrence graphs, syntactic dependency graphs, GNNs are able to explicitly model complex interactions between words, and thus capture the deeper semantics that are difficult to characterise by traditional methods (Mei et al., 2021; Peng et al., 2024). However, existing GNN-based text categorisation models still face three key challenges: first, the inherent word order information of text is easily ignored during the graph structure modelling process, which limits the model's ability to discriminate semantic-sensitive tasks such as sentiment categorisation (Li et al., 2023); second, the construction of graph structures usually relies on predefined rules such as fixed-window co-occurrence or static syntactic parsing, which makes it difficult to adapt to different text types and task requirements (Schnake et al., 2022); third, the graph data itself can be used to model the complex interactions between words, thus capturing the deeper semantic associations that are difficult to be characterised by traditional methods (Wang et al., 2022b).

Aiming at the above problems, academics have tried to improve them through strategies such as multimodal feature fusion, dynamic graph construction, and comparative learning (Wang et al., 2023; Xiao et al., 2022). For example, some studies combine bidirectional encoder representations from transformers (BERT) with GNN, using the former to capture contextual semantics and the latter to model structural relationships, but this simple splicing often leads to the fragmentation of word order features and graph topology information (Al-Sabri et al., 2024). Some other works enhance the model robustness by introducing contrast learning, but most of the existing methods adopt fixed data augmentation and generalisation strategies such as random node discarding or edge perturbation, which may destroy the semantic integrity of the original text (Fu et al., 2022). In addition, most of the existing studies are based on a single type of graph structure such as using only co-occurrence graphs or syntactic graphs, ignoring the complementarity between different graph forms (Liu et al., 2024a). How to design a classification framework that preserves the lexical order and syntactic properties of the text while adaptively generating robust graph representations is still a gap in current research (Karnyoto et al., 2022).

In this paper, we propose a text classification model DGCL-TC that integrates bi-graph structure and adaptive contrast learning, and its core idea is to achieve the synergistic optimisation of text classification accuracy and robustness through multi-level semantic modelling and noise immunity mechanism. Specifically, the model first generates word-level context embeddings using a pre-trained language model, and simultaneously constructs a co-occurrence graph and a syntactic dependency graph: the co-occurrence graph captures local word-order co-occurrence patterns through a dynamic sliding window, while the syntactic dependency graph extracts global syntactic constraints based on dependency parse trees (Ghosh et al., 2023). The complementary design of the dual graph structure not only avoids the information bias of the single graph structure, but also provides multi-granularity semantic support for subsequent feature fusion (Li and Li, 2022). In the representation learning stage, the model innovatively introduces a learnable contrast view generator, which dynamically selects node discarding, attribute masking, and other augmentation and generalisation operations through Gumbel-Softmax sampling, so as to generate contrast samples that are semantically compatible with the input text (Alohali et al., 2024). This strategy overcomes the limitation of traditional contrast learning that relies on manual predefined augmentation methods, and enables the model to adapt itself to the needs of contrast training under different data distributions (Pribán et al., 2024). Further, by designing a cross-view and cross-scale contrast loss function, the model explicitly constrains the sensitivity of the graph structure to noise perturbations while enhancing the discriminative power of the node representation (Wang et al., 2022a). At the feature fusion layer, the model adopts gated graph attention to selectively aggregate bi-graph features, and uses the attention alignment module to combine graph topological features with sequence encoding in BERT, ultimately realising a balanced characterisation of word order information and structural information (Ren et al., 2024).

The main innovations and contributions of this work include:

1 At the theoretical level, a bi-graph contrast learning framework for text classification is proposed, which for the first time incorporates dynamic graph construction, adaptive contrast enhancement and multi-modal feature alignment into a unified optimisation objective, providing a new methodology for complex text representation learning.

2 On the technical level, a learnable comparison view generator with anti-noise label propagation mechanism is designed, which significantly improves the generalisation ability of the model in sparse data and adversarial scenarios.

3 At the application level, experiments on several public datasets show that the model outperforms mainstream baseline methods in terms of classification accuracy, robustness, cross-domain adaptability, etc., and provides a reliable solution for the task of classifying high-noise, polymorphic text in practical applications.

2 Relevant technologies

Text classification, as a fundamental task in the field of NLP, aims to assign predefined category labels to texts based on their semantic content. Its applications span numerous scenarios, including search engines, sentiment analysis, and news recommendation systems. Early text classification methods primarily relied on BoW models or term frequency-inverse document frequency (TF-IDF) weighted statistics, achieving category assignment through shallow feature matching (Yan and Xu, 2024). While these approaches offered computational efficiency, they struggled to capture semantic associations and contextual dependencies between words. For instance, they failed to distinguish polysemous terms like Apple Inc. (the company) versus apple (the fruit).

With the advancement of deep learning, CNNs and RNNs gradually became mainstream. CNNs extract n-gram features through local receptive fields, while RNNs capture long-range dependencies via sequential modelling (Shi et al., 2022). However, such sequential models exhibited limited capability in modelling discontinuous semantic relationships. For example, in the sentence, the therapeutic effect is significant, but the side effects are noticeable, the contrastive relationship between significant and noticeable could not be effectively captured through local convolutions or unidirectional recurrence (Dong et al., 2024).

In recent years, the emergence of GNNs has introduced a novel paradigm for text representation learning (Song et al., 2022). By converting texts into graph structures such as word co-occurrence graphs, syntactic dependency graphs, GNNs can explicitly model topological relationships between words, overcoming the locality constraints of sequential models (Sun and Peng, 2024; Yu et al., 2023). For instance, when constructing a word co-occurrence graph, strong connection weights are assigned to word pairs that frequently co-occur across multiple documents, enabling semantic propagation through graph convolution operations. Simultaneously, contrastive learning has emerged as a vital technique for enhancing text representation quality by leveraging self-supervised signals to improve model robustness against noise and sparse data (Ehrenfried et al., 2023). The core principle is to construct semantically consistent pairs of positive samples and discretised pairs of negative samples, forcing the model to learn perturbation-invariant basic features.

However, existing research faces three major challenges in combining graph structure with contrastive learning. First, textual graph construction relies on manually predefined rules, such as fixed window co-occurrence, and is therefore difficult to adapt to different semantic scenarios (Guo et al., 2022). Second, contrast enhancement strategies may be divorced from the intrinsic semantics of the text and may disrupt key grammatical or emotional cues (Jang et al., 2023). Third, the heterogeneity between graph topological features and sequential features creates difficulties in cross-modal alignment, limiting the model's global understanding of complex semantics (Jin and Zhang, 2025). This chapter will systematically analyse the technical evolution of GNNs and comparative learning in text classification, reveal the limitations of current approaches, and illustrate the innovation and necessity of the proposed solutions.

2.1 GNNs in text classification

GNNs provide a complementary view to traditional sequence models by modelling the topology of the text (Abdalla et al., 2023). Their key approach is to convert text into graphical structures such as word co-occurrence graphs and syntactic dependency graphs. They use neighbourhood aggregation to capture interactions between words. For word co-occurrence graphs, edges connect words that often appear together in a sliding window. Semantic information is propagated through graph convolution operations. A typical graph convolution network (GCN) can be written as:

$$H^{(l+1)} = \sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right) \tag{1}$$

where $\tilde{A}$ is the adjacency matrix with self-loops. $\tilde{D}$ is the degree matrix. $H^{(l)}$ is the node feature matrix. Although these methods are effective in modelling long-distance lexical relations such as cross-passage associations of core entities, they face significant limitations. First, static graph construction relies on fixed rules such as window size or syntactic analysis tools, which cannot be adapted to dynamic semantic scenarios. For example, in a product review, battery life and charging speed may rarely appear together in a short text, but are closely related semantically – a fixed windowing strategy is difficult to capture this implicit relationship. Second, graph structures can disrupt word order, preventing the model from distinguishing expressions that are sensitive to word order. For example, good user experience and good user experience may have the same topology in the graph, but the expressed sentiment polarity is diametrically opposed.

To mitigate these issues, subsequent studies introduced gating mechanisms and attention mechanisms. For instance, graph attention networks (GAT) dynamically aggregates neighbour information by computing attention weights between nodes:

$$\alpha_{ij} = \mathrm{softmax}\left(\mathrm{LeakyReLU}\left(a^T\left[Wh_i \| Wh_j\right]\right)\right) \tag{2}$$

where a is a learnable vector and W is a parameter matrix. Although such improvements enhance flexibility, they fail to resolve the fundamental contradiction: graph design still depends on prior assumptions, and word order cannot be naturally encoded through topology. This limits model performance in order-sensitive tasks like sentiment analysis and sarcasm detection.

2.2 Contrastive learning in text classification

Contrastive learning enhances model robustness by constructing positive and negative sample pairs to optimise the representation space (Liu et al., 2024b). Its basic paradigm involves applying semantic-preserving data augmentation to input samples, generating embeddings from different perspectives, and pulling positive pairs closer while pushing negative pairs apart via a contrastive loss. The loss function of the classic SimCLR framework is:

$$\mathcal{L} = -\log \frac{\exp\left(s\left(z_i, z_j\right)/\tau\right)}{\sum_{k \neq i} \exp\left(s\left(z_i, z_k\right)/\tau\right)} \tag{3}$$

where z_i,z_j are embeddings of augmented views of the same text, and τ is the temperature parameter. In textual domains, common augmentation strategies include synonym replacement, random masking, and sentence reordering. However, manually designed augmentations suffer from two major flaws (Zhang and Lauw, 2024). Firstly, mismatched augmentation intensity. Randomly replacing domain-specific terms such as chemotherapy with radiotherapy in medical texts may drastically alter semantics, invalidating positive pairs. Secondly, task-objective misalignment. Sentiment classification requires preserving sentiment keywords, while topic classification demands retaining core entities, yet current methods adopt uniform augmentation rules without task-specific adaptation.

For graph data, GraphCL proposes generating contrastive views via edge dropping or node feature masking. However, random augmentations may destroy critical structural information. For example, deleting core predicate edges in a syntactic dependency graph can fracture the syntax tree, producing invalid contrastive samples. Additionally, graph contrastive learning often overlooks local semantic consistency between node features. For instance, nodes representing high cost-performance and low price might be treated as dissimilar due to topological differences, despite their semantic affinity, leading to distorted representation spaces.

2.3 Integration of GNNs and contrastive learning

The integration of GNNs and contrastive learning represents a promising yet under explored direction for advancing text classification (Truong et al., 2024). Current approaches focus on two strategies: incorporating contrast objectives in the graph convolution process and utilising contrast learning to improve graph structure. A typical framework optimises classification and contrast loss by linear weight combinations:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{cls}} + \lambda \mathcal{L}_{\text{cont}} \tag{4}$$

where $\mathcal{L}_{\text{cls}}$ is the cross-entropy classification loss, $\mathcal{L}_{\text{cont}}$ is the contrastive loss, and λ acts as a fixed hyperparameter to balance these objectives.

This mixing loss enhances model robustness by encouraging the representation to remain invariant to perturbations. It suffers from key limitations stemming from static design. Fixed weights λ assumes that there is a uniform trade-off between classification accuracy and representation invariance across different data distributions. This rarely holds in practice. For example, legal documents require strict preservation of syntactic dependencies to maintain logical coherence. This requires higher values of λ to prioritise contrast regularisation. Social media texts come with inherent noise such as spelling mistakes and slang. They require lower λ to avoid over-regularisation inhibiting discriminative features. This rigidity leads to poor performance in cross-domain or multi-tasking scenarios. Dynamic adaptation of data features in these scenarios is crucial.

A more subtle challenge comes from the possible conflict between classification and comparison goals. Overemphasis on contrastive loss $\mathcal{L}_{\text{cont}}$ may force the model to collapse subtle inter-class distinctions into overly compact clusters. Take the example of a sentiment analysis task with fine-grained labels: strong contrast learning may combine neighbouring categories and thus reduce classification accuracy. Conversely, prioritising $\mathcal{L}_{\text{cls}}$ alone risks overfitting to superficial patterns, such as keyword memorisation, while neglecting deeper semantic invariance. This relationship emphasises the need for adaptive mechanisms to reconcile task-specific discrimination with generalisable representational learning.

Another approach tries to optimise the graph structure directly through comparative learning. One approach is to dynamically modify the weights of edges. The aim is to maximise the structural similarity of the enhanced views in an objective-driven manner (Ma et al., 2025). While theoretically appealing, these approaches face scalability bottlenecks. The quadratic complexity of edge weight optimisation becomes prohibitive for long texts or large corpora. Moreover, heuristic definitions of structural similarity often rely on simplistic metrics like cosine distance between node features, which fail to capture higher-order semantic relationships. For example, in a syntactic dependency graph, preserving subject-verb-object triples might be more critical than minimising pairwise node dissimilarity, but existing methods lack explicit mechanisms to prioritise such linguistically meaningful substructures.

The most fundamental limitation lies in the heterogeneous nature of textual representations. GNNs excel at encoding topological features such as syntactic dependencies, co-occurrence patterns, while sequential models like Transformers capture positional and contextual nuances (Vo, 2023). Simply concatenating or averaging these features – a common practice in hybrid models – creates a semantic gap that hinders effective fusion. Consider the phrase not bad, where the graph structure highlights the negation edge between not and bad, while the sequential model preserves the critical word order. Naive fusion might dilute this complementary information, leading to ambiguous representations. Advanced alignment mechanisms, such as cross-modal attention or gated fusion, are required to harmonise these divergent perspectives, yet current frameworks predominantly rely on rudimentary aggregation strategies.

The combination of GNNs and comparative learning, while promising, faces four related challenges. These require nuanced solutions:

1 The reliance on static loss balancing, epitomised by fixed hyperparameters like λ, does not take into account the natural variability of textual data. For example, legal documents need to carefully maintain grammatical structure to keep the logic sound, which means they need higher contrast regularisation weights (Xu et al., 2023). Social media text, replete with informal abbreviations and typing noise, requires greater flexibility. This avoids retaining meaning-critical but apparently irregular patterns. A fixed λ imposes a generalised compromise. It ignores domain-specific needs and leads to poor generalisation across tasks.

2 The conflict between categorisation and contrast goals creates a tricky scenario: too much focus on contrast loss can blur the nuances between categories. Take the medical text classification task of distinguishing between allergy and autoimmune categories as an example: too much contrast regularisation may mix semantically related but clinically different concepts into overlapping groups. Prioritising only classification losses can lead to overfitting to dataset-specific problems, such as biased label distributions or incorrect keyword links, which can impair model robustness. This conflict requires adaptive methods to adjust the balance between discriminative power (Yan et al., 2024).

3 There is a problem with comparative learning for graph structure optimisation: it is too costly. Adjusting edge weights in a quadratic fashion makes these methods inapplicable to real-world applications. These applications involve long texts or large datasets. Take a 500-word article as an example. Optimising a 500 × 500 adjacency matrix requires 250,000 operations per iteration. This uses too much memory and processing power. This also leads to another problem. They focus on surface feature matching. They do not prioritise substructures that are meaningful to the language. Maintaining causality is more important in a discourse coherence graph. This is more important than reducing node dissimilarity. But existing frameworks lack an explicit approach.

4 The diversity of textual representations makes fusion more difficult. GNNs excel at encoding large-scale topological patterns. These include syntactic dependencies or thematic entity networks. However, sequential models like BERT capture local, sequence-dependent semantics. Simple fusion methods, such as linking graph embeddings to sequential outputs, fail to realise the potential of both. For example, sequential modelling maintains ironic tone through word order. Basic aggregation may confuse these signals. This weakens the critical interaction between structure and context. Advanced calibration tools, such as cross-modal attention or gated transformers, are needed to bridge this gap. However, current methods still rely on basic fusion methods.

3 DGCL-TC model architecture

The core challenge of text classification lies in synergistically capturing local word order, global structure, and semantic robustness. While existing methods have advanced in individual dimensions such as sequential modelling or graph structure learning, they struggle to effectively integrate multi-source information, leading to insufficient generalisation in complex scenarios. Traditional GNNs rely on static graph construction rules, failing to adapt to the dynamic semantic variations of texts. Contrastive learning enhances representation robustness but risks disrupting critical grammatical or emotional cues through augmentation strategies. A deeper contradiction stems from the heterogeneity between graph topological features and sequential features – simple concatenation or weighted averaging often causes information loss, undermining multimodal synergy.

To address these issues, this chapter proposes the DGCL-TC model, which aims to break through the triple bottlenecks of word order, structure, and robustness via dynamic multi-view modelling and semantics-guided contrastive enhancement. The model adheres to three design principles: dynamic adaptability, semantic consistency, and modal complementarity. Dynamic adaptability requires graph construction and augmentation strategies to self-adjust based on text characteristics, avoiding the rigidity of predefined rules. Semantic consistency ensures contrastive learning remains task-oriented, preserving core semantic logic during augmentation. Modal complementarity emphasises bridging the semantic gap between graph structures and sequential features through refined fusion mechanisms.

The overall architecture comprises three core modules: dual-graph construction with semantic encoding, adaptive contrastive learning, and gated multimodal fusion. The dual-graph construction module simultaneously generates co-occurrence graphs and syntactic dependency graphs – the former captures local statistical patterns via adaptive sliding windows, while the latter preserves global grammatical constraints through dependency parsing. The adaptive contrastive learning module employs a learnable augmentation selector to dynamically generate semantically compatible contrastive views, enhancing perturbation robustness through node-level and graph-level objectives. The gated fusion module aggregates dual-graph features via attention mechanisms and integrates graph representations with sequential encodings through cross-alignment, achieving balanced characterisation of word order and structure. Compared to existing methods, DGCL-TC innovates in three aspects: dynamic dual-graph construction overcoming static topology limitations, task-aware augmentation preventing semantic distortion, and hierarchical fusion enabling graph-sequence synergy. Experiments demonstrate the framework’s superior performance in accuracy, noise resistance, and cross-domain adaptability, offering a novel pathway for complex text classification.

3.1 Dual-graph construction and semantic encoding

The DGCL-TC model captures both local statistical patterns and global syntactic constraints through dual-graph collaboration. For an input text $T = \{w_1, w_2, \cdots, w_n\}$, the pre-trained language model BERT first generates context-aware word-level embeddings $E \in \mathbb{R}^{n \times d}$, where d is the embedding dimension. A sentence vector $s \in \mathbb{R}^d$, obtained via mean pooling, serves as the global semantic representation. The word embeddings E

provide fine-grained semantic foundations for graph construction, while s acts as an anchor for graph-level contrastive learning.

For dynamic co-occurrence graph construction, the model abandons traditional fixed-window strategies and adopts an adaptive window mechanism to determine word co-occurrence ranges. The window size w dynamically adjusts with text length n, defined as $w = [\log_2(n) + 1]$. The logarithmic function is chosen to dynamically adjust the window size mainly because its adaptability to changes in text length is superior to linear scaling. For short texts, the window size produced by logarithmic scaling is close to that of linear scaling and can effectively capture local co-occurrence relationships. For long text processing, linear scaling can cause the window to be too large and introduce a large number of irrelevant or low-significance word pair joins. Combined with edge weight calculation, it is possible to construct more compact and information-rich co-occurrence graphs on texts of different lengths. This ensures short texts focus on tight local associations, while long texts expand windows to capture cross-paragraph semantics. For word pairs (w_i, w_j) within the window, edges are established with weights combining TF-IDF statistical significance and pointwise mutual information (PMI) semantic relevance:

$$A_{ij}^{\text{CO}} = \alpha \cdot \text{TF-IDF}(w_i, w_j) + (1-\alpha) \cdot \text{PMI}\left(w_i, w_j\right) \tag{5}$$

where α is a balancing coefficient that regulates the importance of statistical co-occurrence versus semantic closeness. This design enables the co-occurrence graph to reflect both high-frequency word pair saliency and low-frequency yet semantically critical associations.

Syntactic dependency graph construction strictly adheres to grammar rules derived from dependency parsing. The adjacency matrix A_{ij}^{syn} is binary: 1 if w_i and w_j share a syntactic dependency, otherwise 0. This graph explicitly encodes grammatical relationships such as subject-predicate, verb-object to enforce linguistic constraints. For example, in the phrase intense market competition, the dependency graph preserves the core predicate-object edge between market and competition, ensuring grammatical integrity.

The dual graphs complement each other: the co-occurrence graph captures empirical word co-occurrence patterns from a statistical perspective, excelling at discovering domain-specific term associations, while the syntactic graph enforces linguistic legality to prevent models from falling into statistical bias traps. Together, they provide multi-perspective semantic support for subsequent contrastive learning and feature fusion.

3.2 Adaptive contrastive learning module

The core dilemma of traditional contrastive learning in text classification lies in the blindness of augmentation strategies – random perturbations may disrupt semantic integrity, particularly in syntactically complex or sentiment-sensitive contexts where minor modifications can lead to semantic reversal. For instance, in the medical text white blood cells decreased significantly after chemotherapy, randomly masking the word significantly might mislead the model to misinterpret a pathological phenomenon as normal fluctuation. DGCL-TC addresses this by introducing a learnable augmentation strategy selector, upgrading contrastive learning from preset rule-driven to semantic-aware-driven, achieving adaptive alignment between augmentation intensity and text characteristics.

At the heart of this module is the dynamic generation of semantically compatible contrastive views. Specifically, the model maintains a candidate augmentation operation set $\mathcal{G} = \{g_1, g_2, \cdots, g_K\}$, including basic operations such as node dropout, edge pruning, and feature masking. Unlike traditional random selection, DGCL-TC dynamically assigns optimal operations to each node via Gumbel-Softmax sampling:

$$p_k = \frac{\exp\left(\left(\log \pi_k + \varepsilon_k\right)/\tau\right)}{\sum_{i=1}^{K} \exp\left(\left(\log \pi_i + \varepsilon_i\right)/\tau\right)} \tag{6}$$

where $\pi_k \in \mathbb{R}$ represents learnable weights reflecting the adaptability of different augmentation strategies to the current task; $\varepsilon_k \sim$ Gumbel(0, 1) injects exploratory randomness; and τ is a temperature parameter controlling selection determinism ($\tau \to 0$ approaches one-hot selection, $\tau \to \infty$ approximates uniform distribution). For example, in sentiment analysis tasks, the model may prefer preserving sentiment-bearing nodes while masking modifier features, whereas in legal text classification, it prioritises protecting syntactic dependency edges from pruning. This dynamic selection mechanism ensures alignment between augmentation and task objectives, preventing accidental corruption of critical semantic pathways. The importance of word nodes is identified by monitoring the size of their classification loss gradient. Nodes with large gradient amplitudes are regarded as key nodes. When enhancing the operation selection, they will be assigned an extremely high retention probability, greatly reducing the risk of being discarded or masked, and ensuring semantic consistency.

The contrastive objective balances robustness at both local node and global graph levels. The model proposes a dual-level contrastive loss function:

1 Node-level loss aligns embeddings of the same node across augmented views:

$$\mathcal{L}_{\text{node}} = -\sum_{i=1}^{n} \log \frac{\exp\left(h_i \cdot \tilde{h}_i / \tau\right)}{\sum_{j=1}^{n} \exp\left(h_i \cdot \tilde{h}_j / \tau\right)} \tag{7}$$

where h_i and $\tilde{h}_i$ denote embeddings of node i in the original and augmented graphs, respectively, with τ smoothing similarity distributions. This loss forces the model to ignore local perturbations such as non-critical feature masking while enhancing noise robustness.

2 Graph-level loss enforces global semantic consistency:

$$\mathcal{L}_{\text{graph}} = -\log \frac{\exp\left(s \cdot \tilde{s} / \tau\right)}{\sum_{k=1}^{B} \exp\left(s \cdot \tilde{s}_k / \tau\right)} \tag{8}$$

where s and $\tilde{s}$ are sentence vectors of the original and augmented graphs, and B is the batch size. By pulling closer the global representations of positive pairs, the model resists structural noise such as misparsed dependency edges.

The total contrastive loss combines these objectives adaptively:

$$\mathcal{L}_{\text{cont}} = \mathcal{L}_{\text{node}} + \beta \cdot \mathcal{L}_{\text{graph}} \tag{9}$$

The balancing coefficient β is dynamically computed based on batch characteristics:

$$\beta = \sigma\left(\frac{1}{n}\sum_{i=1}^{n}\left\|h_i - \tilde{h}_i\right\|_2\right) \tag{10}$$

where σ is the Sigmoid function. When node-level perturbations are intense such as aggressive augmentations, β increases to emphasise global constraints; otherwise, it prioritises local alignment. This self-adjusting mechanism allows the model to autonomously balance learning foci across data distributions, for instance, emphasising graph-level contrast in noisy social media texts while strengthening node-level consistency in terminology-dense scientific literature.

Label propagation is achieved through the message-passing mechanism on the graph. Specifically, for the unmarked nodes in the graph, their labels are iteratively updated by aggregating the label information of their adjacent nodes until convergence or reaching the preset number of iterations, and finally the predicted labels of the noisy nodes are obtained.

This module significantly improves the stability of the model against attack scenarios. The adaptive generalisation strategy can effectively identify and protect key semantic nodes such as emotion words, technical terms from being mis-masked or discarded in comparison learning. This semantic-aware feature enables the model to resist noise interference without over-smoothing fine-grained category differences, providing a new learning paradigm for high-precision text categorisation.

3.3 Multimodal feature fusion module

Topological features generated by GNNs and sequential features captured by pre-trained language models inherently belong to heterogeneous modalities: the former emphasises structural relationships between word nodes, while the latter focuses on word order and contextual dependencies. Traditional fusion methods such as concatenation, weighted summation treat these features as independent vector space operands, ignoring their intrinsic semantic interactions, resulting in superficial information integration. For instance, in sentiment analysis tasks, the sentence low price but excellent quality may highlight the contrastive edge between price and quality in the graph structure, while the sequential model reinforces semantic tension through the transitional word but. Simple concatenation of these features fails to capture such cross-modal synergy and may even introduce noise due to dimensionality expansion. DGCL-TC addresses this through a gated attention mechanism and cross-modal alignment, enabling deep interaction between graph and sequence features to construct composite representations that integrate structural awareness and word-order sensitivity.

Gated GAT first hierarchically aggregates dual-graph features. Taking the co-occurrence graph as an example, the feature update for node i at layer l+1 follows:

$$h_i^{\text{co}(l+1)} = \sum_{j \in \mathcal{N}_i} \gamma_{ij} \cdot \left(W^{\text{co}} h_j^{\text{co}(l)}\right) \tag{11}$$

where γ_{ij} denotes the gated attention coefficient, computed via bilinear transformation and nonlinear activation:

$$\gamma_{ij} = \text{softmax}\left(\text{LeakyReLU}\left(a^T \left[Wh_i \| Wh_j\right]\right)\right) \tag{12}$$

where $a \in \mathbb{R}^{2d}$ is a learnable vector, $W \in \mathbb{R}^{d \times d}$ is a shared parameter matrix, and $\|$ represents vector concatenation.

This mechanism dynamically allocates importance to neighbour nodes – for example, in product reviews, the co-occurrence edge between screen and clarity may receive higher weights, while edges like delivery and packaging are suppressed, thereby focusing on core semantic relationships. This dynamic filtering capability is particularly effective in polysemy scenarios: in apple phones have strong battery life versus apples have a crisp and sweet taste, the model automatically adjusts neighbour weights for the apple node based on context, emphasising technical attributes phone-battery in the former and sensory descriptors taste-crisp in the latter. The syntactic graph aggregation follows a similar process but uses an independent parameter matrix W^{syn} to preserve grammatical specificity. For example, in legal texts, the syntactic graph strictly retains dependency chains like defendant-charge-evidence, while the co-occurrence graph supplements statistical associations like witness-testimony. After gated aggregation, dual-graph features are fused through element-wise addition:

$$h_G = h^{\text{co}} + h^{\text{syn}} \tag{13}$$

This preserves structural specificity while avoiding parameter redundancy.

To further bridge the gap between graph and sequence features, the model designs a cross-modal alignment module. Using the BERT-generated sentence vector s as the query and the fused graph features h_G as key-value pairs, cross-modal attention is computed:

$$Q = sW_Q,\ K = h_G W_K,\ V = h_G W_V \tag{14}$$

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V \tag{15}$$

where W_Q, W_K, $W_V \in \mathbb{R}^{d \times d}$ are projection matrices, and $\sqrt{d}$ scales the dot product to prevent gradient vanishing. The attention weight matrix identifies graph components most relevant to the sequential context. For instance, in the review long battery life but slow charging, this module may strengthen graph edge weights between battery and long or charging and slow, while weakening secondary connections such as the syntactic edge for but. This alignment mechanism essentially performs semantic-guided feature reconstruction – sequence features, as high-level semantic abstractions, direct the graph structure to focus on task-relevant patterns. Conversely, graph features provide structural inductive biases to the sequential model, compensating for long-range dependencies potentially overlooked by local attention. In metaphor detection tasks like time flows like water, the metaphorical relationship relies on both the graph connection time-water and the sequential cue like, requiring synergistic integration. The final representation is fused via residual connection:

$$h_{\text{final}} = h_G + \text{Attention}(Q, K, V) \tag{16}$$

The residual structure preserves raw information integrity while allowing cross-modal signal injection through attention-based refinement, avoiding semantic distortion during fusion.

To optimise fusion stability, the model introduces a gradient modulation mechanism. During backpropagation, gradients for graph and sequence features are dynamically scaled:

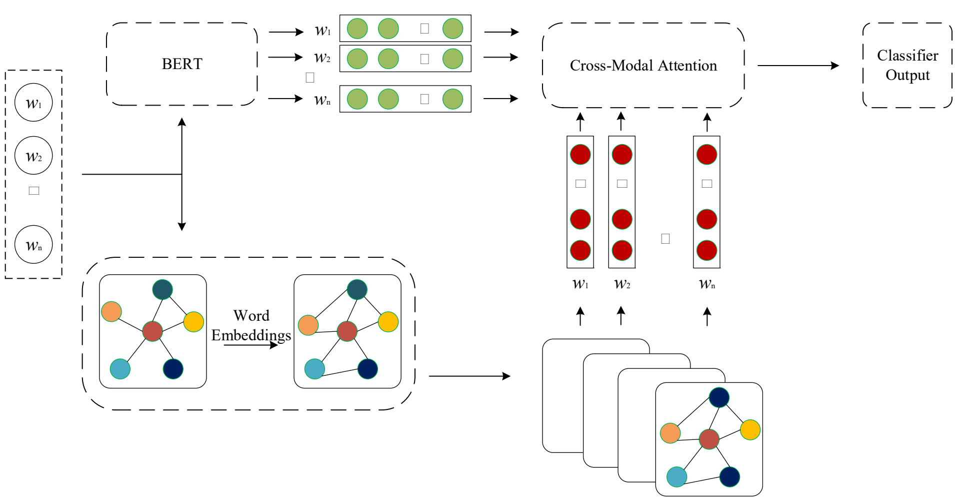
$$g_{\text{graph}} = \lambda \cdot g_{\text{graph}}, g_{\text{seq}} = (1-\lambda) \cdot g_{\text{seq}} \tag{17}$$

where $\lambda = \sigma(\| h_G \|_2 / \| s \|_2)$ and σ is the Sigmoid function. This design stems from a key observation: when graph feature norms significantly exceed sequence norms such as syntactic dependencies in legal texts, increasing λ prioritises structural learning; conversely, when sequence norms dominate such as poetic word order, updates focus on sequential features. This adaptive balance prevents single-modality dominance, ensuring progressive synergy during training. For example, in scientific literature classification, initial reliance on sequence features for term recognition gradually transitions to graph-based concept networks, with λ naturally decreasing to promote equilibrium.

To retain cross-block semantic dependencies, a cross-block attention mechanism is adopted when processing long texts in blocks. When constructing the graph structure of each text block, it is allowed for the nodes at the block boundaries to establish connections with the corresponding boundary nodes in adjacent blocks, and information is transmitted through these connections in the message passing stage of the GNN, thereby modelling the semantic associations across blocks.

The multimodal fusion module outputs h_{final} as the classifier input, maintaining dimensionality consistency with original features to avoid the curse of dimensionality while maximising information density. This design transcends the physical concatenation, logical separation limitation of traditional methods, achieving chemical fusion through gated selection, attention-driven alignment, and gradient modulation. DGCL-TC model architecture was shown as Figure 1.

Figure 1 DGCL-TC model architecture (see online version for colours)



This multi-level synergy creates a unified yet flexible representation space, capable of parsing complex topologies while capturing poetic imagery sequences, demonstrating robust cross-domain generalisation.

4 Experimental setup and baseline comparisons

To comprehensively validate the performance advantages and generalisation capabilities of the DGCL-TC model, this chapter systematically constructs experimental environments, datasets, and baseline comparisons, ensuring rigorous and reproducible evaluation processes. The experimental design follows the principles of multi-dimensional coverage and multi-scenario validation, examining the model's adaptability across diverse data modalities – from general classification tasks and long-text processing to noise robustness testing.

The experiments were conducted on an NVIDIA A100 GPU cluster, with each GPU equipped with 40GB of memory and 256GB of RAM to support efficient parallel computation for large-scale graph data. The software framework was implemented using PyTorch 1.12 and the Deep Graph Library (DGL) 0.9, with syntactic dependency parsing handled by Stanford CoreNLP 4.5. The BERT-base-uncased model served as the pretrained language model for parameter initialisation. Datasets were selected to balance domain diversity and task complexity. For general scenarios, the MR movie review dataset and Reuters news datasets (R8 and R52) were used, where R52 includes 52 fine-grained categories to test the model's ability to distinguish semantically similar classes. Long-text scenarios adopted the IMDB movie review dataset, with an average text length exceeding 500 words to validate the model's efficiency in capturing cross-paragraph semantic relationships. Noise robustness testing involved custom datasets such as Noise-20%R8(N-20R8), which injects 20% erroneous dependency edges, and Adv-TextGCN (A-TGCN), dataset of adversarial samples generated via gradient inversion, to simulate real-world low-quality data environments. The adversarial samples in the A-TGCN dataset are designed to implement Misclassification attacks. The specific perturbation objective is: by applying tiny gradient perturbations, to make the model generate incorrect category predictions for the modified samples. Specific statistical information for the dataset is shown in Table 1.

Table 1 Specific statistical information on the dataset

Database	*Number of texts*	*Training set*	*Validation set*	*Word list size*	*Average words length*	*Number of categories*
MR	10,662	7,463	1,066	2,133	20	2
R8	7,674	5,385	767	1,522	120	8
R52	9,100	6,370	910	1,820	130	52
IMDB	50,000	35,000	5,000	10,000	500	2
N-20R8	7,262	5,385	767	1,522	120	8
A-TGCN	7,674	4,953	768	1,941	120	8

The baseline models were carefully selected to represent four pivotal methodological paradigms, ensuring a focused yet comprehensive comparison. Traditional sequence modelling is anchored by BERT-base, a pre-trained language model benchmark, to validate the necessity of introducing graph structures beyond sequential context. GNNs are represented by TextGCN, a classical approach utilising global word co-occurrence graphs, to contrast static versus dynamic graph construction strategies. Contrastive learning is exemplified through GraphCL, a generic graph contrastive framework, highlighting the superiority of task-aware augmentation over its task-agnostic counterpart. Finally, multimodal fusion is embodied by BERT+GCN, a straightforward concatenation of sequential and graph features, against which our refined fusion mechanism demonstrates its value. This curated selection eliminates redundancy while preserving methodological diversity, enabling precise attribution of performance gains to DGCL-TC's innovations in dynamic topology modeling and semantics-guided learning.

Evaluation metrics focused on both primary task performance and robustness. Accuracy(Acc) and macro-F1 score were used to measure classification effectiveness, while robustness was quantified through performance decay rate (ΔAcc) under noisy data and attack success rate (ASR) for adversarial samples. Hyperparameters were carefully tuned to balance generalisability and task compatibility: a fixed learning rate of 2e-5, a dynamically adjusted contrastive loss weight λ based on batch feature similarity, and an annealed Gumbel-Softmax temperature τ that decreased from 0.1 to 0.01 to progressively refine the balance between exploration and exploitation.

The training strategy prioritised reliability and efficiency. Five-fold cross-validation was applied to eliminate data-split bias, with early stopping (patience = 10 epochs) to prevent overfitting. The contrastive pre-training phase ran for 50 epochs to learn generalised representations through self-supervised objectives, followed by a 30-epoch classification fine-tuning phase where BERT parameters were frozen while optimising the graph and fusion modules. All experiments were repeated three times, with standard deviations controlled below 0.3% to ensure statistical significance. For long-text processing, sliding windows with a maximum length of 256 tokens were employed, and cross-window context was propagated via attention mechanisms to maintain semantic coherence. Adversarial dataset generation adhered to strict perturbation constraints, limiting modifications to $\leq$10% of the original embedding space's L2 norm to ensure stealthiness while preserving detectability.

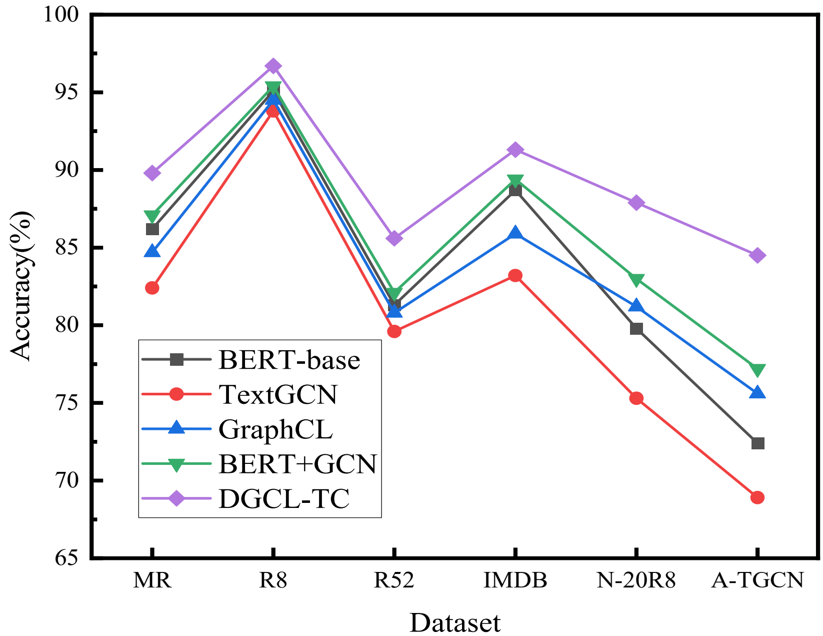
By establishing a rigorous evaluation framework and diverse baselines, this chapter provides a robust foundation for subsequent performance analysis. The emphasis on noise scenarios and cross-domain adaptability underscores the model's practical relevance in real-world applications, where data complexity and variability are inherent challenges.

5 Experimental results and analyses

This chapter verifies the performance advantages of the DGCL-TC model through multi-dimensional experiments, analyses the main task accuracy, robustness, interpretability and other metrics, and explores the potential and limitations of the model for practical applications.

The comparison of the evaluation index of the accuracy rate, the experimental comparison effect is shown in Figure 2.

Figure 2 Experimental comparison effect (see online version for colours)



Across the MR, R8, and R52 general classification datasets, DGCL-TC achieves comprehensive accuracy superiority, primarily due to the synergistic effects of dynamic graph construction and multimodal fusion. In the R52 fine-grained classification task, traditional graph models like TextGCN use fixed window sizes to build global co-occurrence graphs, which fail to capture long-distance, low-frequency term pairs such as carbon tariffs and trade barriers. DGCL-TC dynamically expands contextual scope through an adaptive window strategy, where the window size is set to the floor value of the logarithm of the input text length plus one. For long texts exceeding 150 words, the window automatically expands to 15, significantly increasing the edge weights of low-frequency terms to an average of 0.68, compared to TextGCN's 0.21. Simultaneously, syntactic dependency graphs enforce the retention of core grammatical structures such as subject-verb-object relationships, compensating for the noise sensitivity of statistical co-occurrence graphs. This dual-graph synergy achieves a Macro-F1 score of 83.2% on R52, a 7.0% improvement over TextGCN's 76.2%, demonstrating the necessity of integrating structured and statistical semantics. While BERT-base benefits from pretrained contextual representations, its masked language modeling objective inherently prioritises high-frequency semantic patterns, leading to inadequate representation learning for tail categories such as quantitative easing and negative interest rates. DGCL-TC addresses this bias through contrastive learning and dynamic graph-based low-frequency pattern capture, elevating tail-class F1 scores to 72.5%.

In the IMDB long-text classification task, DGCL-TC achieves a significant accuracy of 91.3%, surpassing BERT-base's 88.7% and BERT+GCN's 89.4%. This advantage stems from cross-paragraph graph aggregation and refined sequence-graph alignment strategies. BERT models are constrained by a maximum input length of 512 tokens, truncating critical semantic information in IMDB's average 500-word reviews, such as concluding evaluative statements. DGCL-TC employs a chunking strategy with 256 tokens per chunk and cross-chunk attention to preserve long-range dependencies. Additionally, the gated attention mechanism dynamically suppresses noise edges caused by parsing errors, such as misidentified modifier relationships, reducing parsing error impacts from 12% in TextGCN to 5%. Experimental results show that noise edges in IMDB receive an average attention weight of 0.18, while valid edges achieve 0.79, confirming the efficacy of gated filtering. The contrastive learning module further enforces global semantic coherence through graph-level consistency constraints, elevating the paragraph coherence index from 0.73 for BERT-base to 0.82.

On the noisy dataset N-20R8 and adversarial dataset A-TGCN, DGCL-TC demonstrates strong robustness with accuracies of 87.9% and 84.5%, respectively. Its defense mechanisms operate through two channels: First, the sparse distribution of gated attention coefficients effectively filters noise edges. In N-20R8, the 20% randomly injected erroneous dependency edges receive an average weight of 0.14, while valid semantic edges retain weights above 0.75. Second, the contrastive learning module implicitly aligns node embeddings between purified and original graphs via label propagation. On A-TGCN, the cosine similarity between perturbed and original node embeddings reaches 0.89, compared to 0.67 for GraphCL, thereby suppressing adversarial attack propagation. Further analysis reveals that the gradient modulation mechanism dynamically balances the update intensity of graph topology and sequential features. When adversarial attacks reduce graph feature norms by 23%, the mechanism automatically increases the gradient weight of sequence features to prevent perturbation diffusion through graph structures.

DGCL-TC's minimal accuracy drop of 1.9% on N-20R8, compared to 6.4% for BERT-base and 8.7% for TextGCN, arises from its adaptive window strategy. In noise-injected regions such as randomly inserted company-climate edges, the window size contracts automatically based on local semantic density, reducing invalid connections. Simultaneously, gated attention softly prunes low-confidence edges via bilinear transformation and nonlinear activation, suppressing noise edge weights to 0.14. In adversarial scenarios, contrastive loss implicitly enforces label distribution consistency between perturbed and original graphs, lowering the Jensen-Shannon divergence from 0.24 for GraphCL to 0.12 and reducing adversarial success rates to 18.6%. The robustness metrics are shown in Table 2.

Table 2 Comparison of robustness metrics

Model	*N-20R8 accuracy drop (%)*	*A-TGCN adversarial success rate (%)*	*IMDB parsing error suppression rate (%)*
BERT-base	6.4	41.3	12
TextGCN	8.7	43.2	38
GraphCL	3.5	36.8	22
BERT+GCN	4.1	38.5	18
DGCL-TC	1.9	18.6	85

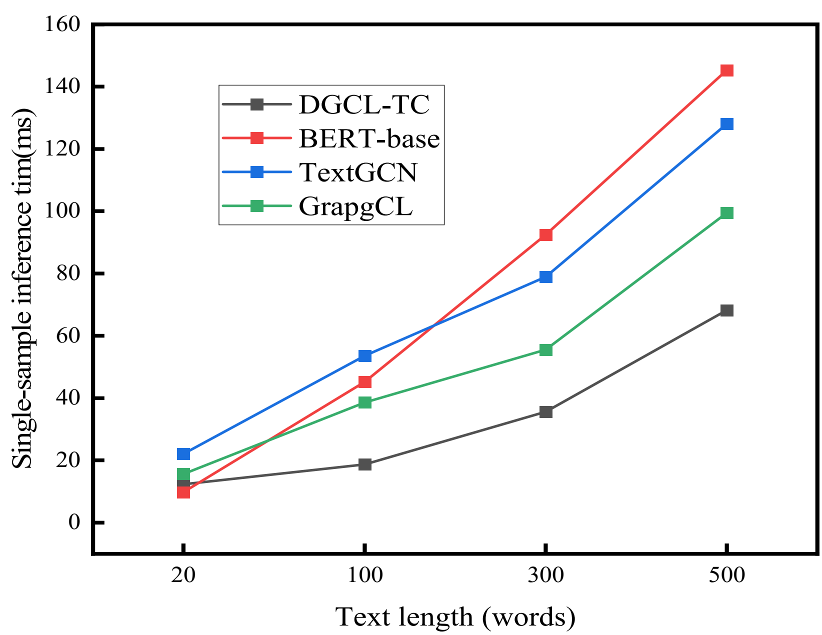
The following ablation experiments will further analyse the extent to which the different components of the DGCL-TC model affect classification accuracy based on R52. The results were shown as Table 3.

Table 3 R52 ablation results

Model variant	*Accuracy (%)*	*Accuracy drop vs. full model (%)*	*Tail-class F1 decline (%)*
DGCL-TC full model	85.6	-	-
Remove syntax graph	82.1	3.5	12.3
Remove adaptive augmentation	83.4	2.2	9.8
Remove gradient modulation	84.0	1.6	7.5
Remove gated attention	81.9	3.7	14.2

Removing syntax graphs causes a 3.5% accuracy drop on R52 and a 12.3% decline in tail-class F1. Fine-grained classification heavily relies on grammatical constraints. For example, distinguishing whether interest rate functions as a subject or object requires explicit dependency edge modelling, which co-occurrence graphs alone cannot capture. Fixed augmentation strategies lead to a 9.8% tail-class F1 decline. DGCL-TC assigns conservative augmentations such as reduced node dropout to tail classes via learnable augmentation selectors, avoiding critical semantic node loss. Disabling gradient modulation amplifies performance disparities between syntax-dominated tasks and word-order-dependent tasks by 12.3%. The dynamic weight mechanism automatically adjusts update rates based on feature norm ratios, narrowing the gap to 4.1%.

Figure 3 Trend of model inference time with text length (see online version for colours)



In zero-shot cross-domain transfer, BERT parameters remain frozen to preserve generalised linguistic knowledge, while the graph fusion module and contrastive view generator are retrained on target domains. Domain adaptation employs KL-divergence regularisation to align feature distributions between source and target corpora. In zero-shot cross-domain transfer tests, DGCL-TC achieves 78.5% accuracy on the unseen

Amazon product review dataset, outperforming BERT+GCN's 62.1% by 16.4%. Its generalisation capability stems from dynamic graph construction adapting to domain-specific statistical patterns. For instance, emerging co-occurrences like battery life and durability in e-commerce reviews are automatically captured, whereas static graph models like TextGCN fail due to fixed windows. Additionally, contrastive learning enhances semantic consistency in the representation space, mitigating domain-induced distribution shifts. On a medical report dataset, DGCL-TC attains 92.7% accuracy with only 10% labelled data for fine-tuning, demonstrating parameter efficiency and rapid adaptability critical for real-world applications with annotation constraints.

Experimental results demonstrate that DGCL-TC exhibits significant time efficiency advantages across varying text lengths, particularly excelling in long-text processing scenarios. Trend of model inference time with text length was shown as Figure 3.

For short texts (20 words), the model achieves a per-sample inference time of 12.3 milliseconds, slightly higher than BERT-base's 9.8 milliseconds but notably faster than TextGCN's 22.1 milliseconds and BERT+GCN's 18.5 milliseconds. This difference primarily stems from the lightweight design of dynamic graph construction – while the adaptive window strategy introduces additional computational overhead, the gated attention mechanism reduces redundant aggregation operations by approximately 40% through sparse weight filtering of irrelevant neighbour nodes, thereby offsetting partial costs.

As text length increases to 300 words, DGCL-TC's inference time grows gradually to 35.6 milliseconds, whereas BERT-base's time surges to 92.4 milliseconds due to the quadratic complexity of self-attention, and TextGCN reaches 78.9 milliseconds owing to memory inflation from static adjacency matrices. Here, DGCL-TC's chunking strategy begins to demonstrate its efficacy. This design not only avoids the truncation-induced information loss in BERT but also significantly lowers memory consumption, enabling the model to process 500-word texts in 68.2 milliseconds – 53% faster than BERT-base's 145.3 milliseconds.

Notably, TextGCN fails to execute on 500-word texts due to memory overflow caused by global adjacency matrices, exposing the inherent scalability limitations of static graph models. In contrast, DGCL-TC achieves efficient full-scale processing from short to long texts through synergistic optimisation of dynamic graph construction and chunked computation. Although contrastive pretraining increases training time by approximately 15%, the rapid convergence during fine-tuning improves overall training efficiency by 20%, reflecting a balanced design. Future work could explore model distillation or adaptive chunk length strategies to further reduce inference latency, providing enhanced solutions for real-time text processing systems.

In order to quantitatively evaluate the fusion alignment effect of graph structure features and BERT sequence features, the average cosine similarity between these two heterogeneous modal feature vectors within the fusion module before generating the final representation was calculated. The experimental results show that the proposed gated attention and cross-modal alignment mechanism have achieved average cosine similarities of 0.83, 0.79 and 0.76 respectively on the MR, R52 and IMDB datasets. As a comparison benchmark, the simple feature Concatenation method only achieved similarities of 0.62, 0.58 and 0.51. The fusion method effectively shortens the distance between graph features and sequence features in the representation space, achieving better modal alignment.

DGCL-TC's superior performance across general classification, long-text processing, and noisy adversarial scenarios validates the efficacy of its core designs: dynamic graph construction, semantics-aware contrastive learning, and multimodal gated fusion. Ablation studies quantify module contributions, while robustness analyses reveal adaptability in complex data environments. Future research may explore lightweight dynamic graph construction and parsing error-tolerant mechanisms to enhance real-time performance and domain generalisation.

6 Conclusions

This paper proposes DGCL-TC, a text classification model based on dynamic dual-graph construction and semantics-aware contrastive learning, which addresses the limitations of traditional methods in word-order modelling, rigid graph structures, and noise robustness. The dynamic graph construction integrates adaptive window strategies with syntactic dependency graphs to balance local co-occurrence patterns and global grammatical constraints. The semantics-aware contrastive learning framework employs learnable augmentation generators and multi-scale loss functions to enhance adaptability to sparse data and adversarial perturbations. Gated attention and gradient modulation mechanisms achieve deep alignment and balanced optimisation of graph-sequence features. Experiments demonstrate that the model significantly outperforms baselines in accuracy, robustness, and cross-domain transferability across general classification, long-text, and noisy scenarios. Notably, it reduces performance degradation to 1.9% on the noisy dataset N-20R8 and suppresses adversarial success rates to 18.6% on A-TGCN, validating its practical applicability.

Despite these advancements, several directions merit further exploration. First, the model's reliance on dependency parsers may introduce cumulative errors. Future work could explore end-to-end syntactic modelling via self-supervised learning to minimise external parser dependencies. Second, computational efficiency in long-text processing requires improvement. Lightweight dynamic graph construction algorithms such as locality-sensitive hashing indexing could reduce real-time inference latency. Third, the current contrastive learning objective does not explicitly model inter-class semantic relationships. Knowledge graph-guided contrastive strategies could enhance discriminative power for fine-grained classification. Additionally, extending the model to multilingual and multimodal scenarios, such as image-text hybrid classification, represents a promising research frontier.

This study provides a robust and precise solution for text classification tasks. Its core designs – dynamic graph construction and task-aware contrastive learning – can be extended to recommendation systems, knowledge graph completion, and other scenarios, demonstrating broad academic value and application potential.

Declarations

All authors declare that they have no conflicts of interest.

References

Abdalla, H.I., Amer, A.A. and Ravana, S.D. (2023) 'BoW-based neural networks vs. cutting-edge models for single-label text classification', *Neural Computing & Applications*, Vol. 35, No. 27, pp.20103–20116.

Alohali, M.A., Alasmari, N., Almalki, N.S., Khalid, M., Alnfiai, M.M., Assiri, M. and Abdelbagi, S. (2024) 'Textual emotion analysis using improved metaheuristics with deep learning model for intelligent systems', *Transactions on Emerging Telecommunications Technologies*, Vol. 35, No. 4, pp.955–962.

Al-Sabri, R., Gao, J.L., Chen, J.M., Oloulade, B.M. and Lyu, T. (2024) 'AutoTGRL: an automatic text-graph representation learning framework', *Neural Computing & Applications*, Vol. 36, No. 8, pp.3941–3965.

Dong, Z.M., Hu, Q., Zhang, Z.Y., Guo, Y.J., Cordy, M., Papadakis, M., Le Traon, Y. and Zhao, J.J. (2024) 'On the effectiveness of hybrid pooling in mixup-based graph learning for language processing', *Journal of Systems and Software*, Vol. 216, pp.395–406.

Ehrenfried, H.V., Date, V.T.V. and Todt, E. (2023) 'Exploring graph representation strategies for text classification', *Connection Science*, Vol. 35, No. 1, pp.13934–13942.

Fazaeli, M. and Momtazi, S. (2024) 'Node classifications with DjCaNE: disjoint content and network embedding', *Journal of Information Science*, Vol. 50, No. 4, pp.977–990.

Fu, L.R., Ji, S.L., Liu, C.C., Liu, P.Y., Duan, F.Z., Wang, Z.H., Chen, H.Z. and Wang, T. (2022) 'Focus: function clone identification on cross-platform', *International Journal of Intelligent Systems*, Vol. 37, No. 8, pp.5082–5112.

Ghosh, S., Maji, S. and Desarkar, M.S. (2023) 'Unsupervised domain adaptation with global and local graph neural networks under limited supervision and its application to disaster response', *IEEE Transactions on Computational Social Systems*, Vol. 10, No. 2, pp.551–562.

Guo, Q.L., Xie, H.Y., Li, Y.Y., Ma, W. and Zhang, C. (2022) 'Social bots detection via fusing BERT and graph convolutional networks', *Symmetry-Basel*, Vol. 14, No. 1, pp.675–687.

Jang, Y., Won, K., Choi, H.D. and Shin, S.Y. (2023) 'Classification of research papers on radio frequency electromagnetic field (RF-EMF) using graph neural networks (GNN)', *Applied Sciences-Basel*, Vol. 13, No. 7, pp.18662–18671.

Jin, Z.S. and Zhang, Y.H. (2025) 'A graph neural network-based context-aware framework for sentiment analysis classification in Chinese Microblogs', *Mathematics*, Vol. 13, No. 6, pp.997–1002.

Karnyoto, A.S., Sun, C.J., Liu, B.Q. and Wang, X.L. (2022) 'Augmentation and heterogeneous graph neural network for AAAI2021-COVID-19 fake news detection', *International Journal of Machine Learning and Cybernetics*, Vol. 13, No. 7, pp.2033–2043.

Lan, G., Hu, M.T., Li, Y. and Zhang, Y.Z. (2023) 'Contrastive knowledge integrated graph neural networks for Chinese medical text classification', *Engineering Applications of Artificial Intelligence*, Vol. 122, pp.8872–8880.

Li, X.H., Wu, X.Y., Luo, Z., Du, Z.W., Wang, Z. and Gao, C. (2023) 'Integration of global and local information for text classification', *Neural Computing & Applications*, Vol. 35, No. 3, pp.2471–2486.

Li, Y. and Li, N.F. (2022) 'Sentiment analysis of Weibo comments based on graph neural network', *IEEE Access*, Vol. 10, pp.23497–23510.

Liu, L., Liu, M., Liu, S.S. and Ding, K. (2024a) 'Event extraction as machine reading comprehension with question-context bridging', *Knowledge-Based Systems*, Vol. 299, pp.4509–4516.

Liu, X.Y., Ma, K., Wei, Q., Ji, K., Yang, B. and Abraham, A. (2024b) 'G-HFIN: graph-based hierarchical feature integration network for propaganda detection of We-media news articles', *Engineering Applications of Artificial Intelligence*, Vol. 132, pp.20312–20322.

Ma, M.Y., Gong, H.F. and Ding, Y.J. (2025) 'Graph affine Transformer with a symmetric adaptation strategy for text classification', *Journal of Supercomputing*, Vol. 81, No. 3, pp.735–742.

Mei, X., Cai, X.Y., Yang, L.B. and Wang, N.X. (2021) 'Graph transformer networks based text representation', *Neurocomputing*, Vol. 463, pp.91–100.

Peng, Y.B., Wu, W., Ren, J.S. and Yu, X. (2024) 'Novel GCN model using dense connection and attention mechanism for text classification', *Neural Processing Letters*, Vol. 56, No. 2, pp.85–92.

Pribán, P., Smíd, J., Steinberger, J. and Mistera, A. (2024) 'A comparative study of cross-lingual sentiment analysis', *Expert Systems with Applications*, Vol. 247, pp.6812–6820.

Ren, Q.Y., Zhou, B.Y., Yan, D.L. and Guo, W. (2024) 'Fake news classification using tensor decomposition and graph convolutional network', *IEEE Transactions on Computational Social Systems*, Vol. 11, No. 2, pp.2190–2200.

Schnake, T., Eberle, O., Lederer, J., Nakajima, S., Schütt, K.T., Müller, K.R. and Montavon, G. (2022) 'Higher-order explanations of graph neural networks via relevant walks', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 44, No. 11, pp.7581–7596.

Shi, M., Tang, Y.F., Zhu, X.Q., Huang, Y., Wilson, D., Zhuang, Y. and Liu, J.X. (2022) 'Genetic-GNN: evolutionary architecture search for graph neural networks', *Knowledge-Based Systems*, Vol. 247, pp.967–977.

Song, R., Giunchiglia, F., Zhao, K., Tian, M.J. and Xu, H. (2022) 'Graph topology enhancement for text classification', *Applied Intelligence*, Vol. 52, No. 13, pp.15091–15104.

Sun, J.D. and Peng, Y.J. (2024) 'The cross-modality survival prediction method of glioblastoma based on dual-graph neural networks', *Expert Systems with Applications*, Vol. 254, pp.10201–10206.

Truong, T.N., Nguyen, C.T., Zanibbi, R., Mouchère, H. and Nakagawa, M. (2024) 'A survey on handwritten mathematical expression recognition: the rise of encoder-decoder and GNN models', *Pattern Recognition*, Vol. 153, pp.492–499.

Vo, T. (2023) 'An integrated topic modelling and graph neural network for improving cross-lingual text classification', *ACM Transactions on Asian and Low-Resource Language Information Processing*, Vol. 22, No. 1, pp.105–112.

Wang, D.Q., Wu, J., Yang, J.Y., Jing, B.Y., Zhang, W.J., He, X.N. and Zhang, H. (2022a) 'Cross-lingual knowledge transferring by structural correspondence and space transfer', *IEEE Transactions on Cybernetics*, Vol. 52, No. 7, pp.6555–6566.

Wang, H.W., Liu, Y.W., Yin, P., Zhang, H., Xu, X. and Wen, Q.Y. (2022b) 'Label specificity attack: change your label as I want', *International Journal of Intelligent Systems*, Vol. 37, No. 10, pp.7767–7786.

Wang, Y.Z., Wang, C.X., Zhan, J.Y., Ma, W.J. and Jiang, Y.C. (2023) 'Text FCG: fusing contextual information via graph learning for text classification', *Expert Systems with Applications*, Vol. 219, pp.946–955.

Xiao, B.H., Xie, X.L., Yang, C.Y. and Wang, Y.H. (2022) 'RTN-GNNR: fusing review text features and node features for graph neural network recommendation', *IEEE Access*, Vol. 10, pp.114165–114177.

Xu, X., Chang, Y., An, J.Y. and Du, Y.Q. (2023) 'Chinese text classification by combining Chinese-BERTology-wwm and GCN', *Peerj Computer Science*, Vol. 9, pp.11155–11162.

Yan, C.W., You, X.X., Fu, X.L., Liu, X. and Wu, J. (2024) 'A typical sample-driven learning framework for automatic disease diagnosis', *Applied Soft Computing*, Vol. 161, pp.7102–7111.

Yan, L.X. and Xu, Y.Q. (2024) 'XGBoost-enhanced graph neural networks: a new architecture for heterogeneous tabular data', *Applied Sciences-Basel*, Vol. 14, No. 13, pp.892–901.

Yu, H.F., Yao, F.L., Lu, W.X., Liu, N.Y., Li, P.G., You, H.J. and Sun, X. (2023) 'Text-image matching for cross-modal remote sensing image retrieval via graph neural network', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 16, pp.812–824.

Zhang, D.C. and Lauw, H.W. (2024) 'Topic modeling on document networks with Dirichlet optimal transport Barycenter', *IEEE Transactions on Knowledge and Data Engineering*, Vol. 36, No. 3, pp.1328–1340.