Uncertainty-Aware Contrastive Sentence Embedding With Local Context Representation for Text Classification

Han Liu[1,2], Jiaqing Zhan[1], Zhichao Chen[1], Qin Zhang[1]*
[1]College of Computer Science and Software Engineering, Shenzhen University
[2]Guangdong Provincial Key Laboratory of Intelligent Information Processing, Shenzhen University
{han.liu,qinzhang}@szu.edu.cn
{zhanjiaqing2023,2110276168}@email.szu.edu.cn

Abstract

In real-world applications of natural language processing, it is essential to effectively adapt a pre-trained model to a downstream task. While text classification is undertaken as a downstream task, it is crucial to produce meaningful sentence embedding that is adaptive to the task. In this paper, we explore how to effectively adapt a pre-trained model for extracting meaningful context representations from sentences, and propose an uncertainty-aware contrastive sentence embedding approach that involves addressing language ambiguity and inter-class separability for a text classification task. Specifically, we design an end-to-end strategy for driving the process of learning to transform a word embedding matrix into a contextualized sentence vector and to quantify the representation uncertainty of the sentence, while the word embedding matrix is produced by a pre-trained model without fine-tuning, and a label-wise contrastive learning strategy is designed to enhance intra-class compactness and inter-class separability. The results on public data sets show that a considerable improvement of text classification accuracy is achieved by adopting the proposed approach in comparison with using those state-of-the-art methods[1].

1 Introduction

In recent years, there have been significant advances in natural language processing (NLP), since the emergence of Transformer-based models such as bidirectional encoder representations from transformers (BERT) (Devlin et al., 2018) and generative pre-trained transformers (GPT) (Radford et al., 2018). Specifically, BERT is a kind of encoder-only transformers, which are typically adopted for natural language understanding, whereas GPT is a type of decoder-only transformers, which aim at natural language generation.

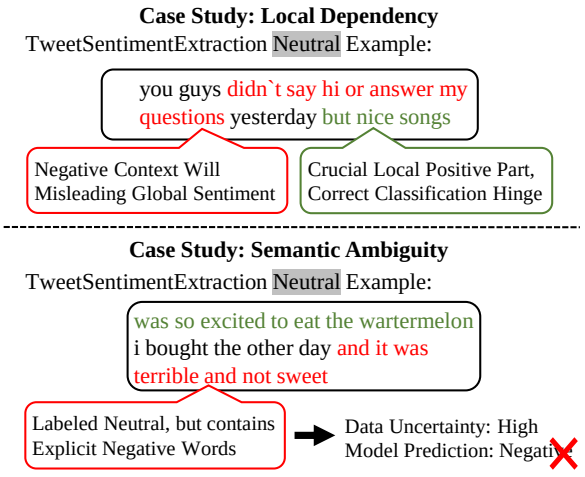

Figure 1: Case study about local dependency and semantic ambiguity.

In real-world applications, it has been quite common to apply pre-trained language models to undertake various downstream tasks in the setting of transfer learning. Moreover, sentence embedding is quite important for some special types of downstream tasks such as text classification (Ni et al., 2022; Gao et al., 2021; Kumar and Raman, 2022) and semantic textual similarity (Reimers and Gurevych, 2019; Gui and Xiao, 2024). In this context, it is essential to adapt a pre-trained model into each specific downstream task and produce meaningful sentence embedding when necessary.

Those state of the art embedding models are typically decoder-only transformers, which work by taking a causal attention mechanism to learn the embedding of each token through exploiting information from all of the preceding tokens in a text sequence and finally taking the EOS token embedding to represent the sentence. For natural language understanding tasks, encoder-only transformers have been adopted popularly as embedding models, which work by taking a bidirectional self-attention mechanism to learn the embedding of each token through exploiting all of the tokens in

*Corresponding author
[1]https://github.com/stickDONTdip/Uncertainty-Aware-Contrastive-Sentence-Embedding

Findings of the Association for Computational Linguistics: ACL 2026, pages 38702–38716

a text sequence and finally taking the CLS token embedding to represent the sentence.

The above strategies of sentence embedding can work well on global representation. However, as shown in Figure 1, a sentence may involve strong dependencies among only those tokens that are close to each other, or exhibit semantic ambiguity where explicit negative words appear in a neutral context. Therefore, in addition to global representation of a sentence, it is necessary to learn local context representations and address semantic uncertainty to enhance sentence embedding.

Moreover, existing methods of sentence embedding mostly work in a deterministic manner. However, some sentences may present very certain meaning whereas some others may be ambiguous as shown in Figure 1. A more detailed case analysis is provided in A.4. From this point of view, it is necessary to incorporate uncertainty learning into sentence embedding, such that the representation uncertainty of each sentence can be quantified to reflect the difficulty of understanding the sentence and less attention is paid to learning embeddings of ambiguous sentences to avoid overfitting. In other words, while involving local context learning can sufficiently enhance the language understandability of an embedding model, representation uncertainty is considered to originate mainly from text ambiguity and uncertainty learning is undertaken to help reduce the impacts of ambiguous sentences.

In this paper, we choose text classification as a downstream task and explore how to effectively adapt a pre-trained model for extracting meaningful context representations from sentences. Specifically, we propose an uncertainty-aware representation learning approach to achieve sentence embedding that is adaptive to a text classification task. This paper has the following contributions:

- We have proposed an uncertainty-aware sentence embedding approach, which involves learning to transform a word embedding matrix into a sentence vector and to quantify representation uncertainty of the sentence.
- We have designed a strategy of extracting multi-grained local context representation to promote the uncertainty learning module to better estimate the clarity of the semantic meaning of an input sentence, where our label-wise contrastive learning strategy is designed to enhance intra-class compactness and inter-class separability in embedding space.
- The experimental results indicate that an improvement of text classification accuracy is achieved by adopting our proposed approach in comparison with using those state of the art methods and our approach can help smaller embedding models outperform larger models.

2 Related Works

Text embedding learning is a fundamental task in NLP, aiming to extract features from text. Early text embedding methods, such as Word2Vec (Mikolov et al., 2013) and GloVe (Pennington et al., 2014), rely on averaging static word vectors, thereby failing to capture contextual nuances. Although pre-trained models like BERT (Devlin et al., 2018) and RoBERTa (Liu et al., 2019) introduce contextualized representations, their direct sentence embeddings often exhibit low quality, frequently underperforming static baselines in semantic textual similarity tasks.

The increasing number and improved quality of diverse text datasets across various tasks for large language models (LLMs), such as GTE (Li et al., 2023; Zhang et al., 2024), NVEmbed-v2 (Lee et al., 2025a), and GritLM-7B (Muennighoff et al., 2025), have enabled the generation of higher-quality representations compared to encoder-only methods (Nie et al., 2024). Representative works, including Gemini embedding (Lee et al., 2025b), Qwen3-Embedding (Zhang et al., 2025a), Linq-Embed-Mistral (Choi et al., 2024), and Seed1.5 (Guo et al., 2025), have become the dominant backbones for top-performing text embedding models on the MTEB English benchmark.

However, most generative LLMs, using contrastive learning to optimize the representation space, are still struggling to capture the holistic semantic meaning due to the causal attention mask. To create bidirectional contextualized representations, LLM2Vec (BehnamGhader et al., 2024) converts causal attention to bidirectional one with an all-ones matrix. Similarly, KaLM (Zhao et al., 2025) removes the causal mask to enable fully bidirectional attention and integrates contrastive distillation. Distinctively, Embeddinggemma (Vera et al., 2025) initializes an encoder-only model from the Gemma-3 decoder to naturally leverage full bidirectional context. Echo-mistral (Springer et al., 2024) feeds the input sentence to LLMs twice. The final contextualized embeddings can then be extracted from the second occurrence of the sentence.

To complement global representations and capture local dependencies limited by the causal mask in LLMs, we exploit multi-grained local contexts from the last hidden states. We propose a unified framework that addresses semantic ambiguity via uncertainty learning and enhances inter-class separability through label-wise contrastive learning.

3 Preliminaries

As introduced in Chang et al. (2020); Abdar et al. (2021); Huang et al. (2024), uncertainty can be categorised into two types, namely, data uncertainty and model uncertainty. In particular, the former type of uncertainty captures noise existing in the training data, whereas the latter type captures noise inherent in a model learned from low quality data.

In the field of computer vision (Liu et al., 2025a; Sheng et al., 2025; Wu et al., 2025; Liu et al., 2025b), data uncertainty learning has been considered as a way of improving the effectiveness of feature representation and the prediction performance in some studies (Li et al., 2021; Shen et al., 2023a; Zhang and Lv, 2024; Gupta et al., 2024). As introduced by Chang et al. (2020), if a dataset is corrupted by noise, data uncertainty learning can be set by hypothesizing that each sample x_i has an expected feature representation $f(x_i)$ and its actually obtained representation z_i in the latent space relates to $f(x_i)$ and the input-dependent noise $n(x_i)$, i.e., $z_i = f(x_i) + \epsilon n(x_i)$, where ϵ follows a standard Gaussian distribution.

In this context, the expected feature representation $f(x_i)$ and the uncertainty $n(x_i)$ of each sample x_i in the latent space are learned simultaneously, which can be treated as the mean and variance of a Gaussian distribution that is learned from the training data, while assuming that x_i is added Gaussian noise with mean of zero and input-dependent variance. Moreover, the setting of $z_i = f(x_i) + \epsilon n(x_i)$ is considered as a reparameterization trick to address the issue that obtaining z_i through sampling from the learned Gaussian distribution $\mathcal{N}(f(x_i), n(x_i))$ is a non-differentiable operation (Kingma and Welling, 2014).

4 Method

Building upon the concepts of data uncertainty learning (Chang et al., 2020), we propose a framework for sentence embedding that focuses on modeling the uncertainty of features while learning local contexts. We assume that input sentences naturally possess different levels of semantic ambiguity, which results in varying difficulty for classification tasks. To address this concern, our method estimates the representation uncertainty of each sample by calculating the variance of a learned distribution within the feature space. Specifically, we convert the fixed token embeddings from a pre-trained model into Gaussian distributions. This strategy enables the model to measure confidence of predictions and effectively mitigate the impact of text ambiguity on representation learning.

To refine the feature space geometry (Frosst et al., 2019; Khosla et al., 2020), we design a label-wise contrastive learning strategy. While uncertainty modeling assesses individual sample quality, standard classification objectives often fail to ensure sufficient inter-class separability. By utilizing label information to align samples within the same category, the designed strategy helps enhance both intra-class density and inter-class separability. This combination yields representations that are both uncertainty-aware and discriminative.

The overall procedure of the proposed approach is illustrated in Figure 2. We freeze the pre-trained parameters, restricting optimization exclusively to the lightweight downstream modules. Using the Qwen3-Embedding model as a representative LLM, we first obtain the token embedding matrix for an input sequence x_i. Specifically, hidden states of the last layer serve as the initial feature representation, denoted as mat_i. Similar to recent probabilistic embedding approaches (Shen et al., 2023b; Yoda et al., 2024), this matrix is subsequently mapped to a Gaussian distribution $dist_i$ by our designed uncertainty-aware one-dimensional convolutional neural network, formally denoted as f.

The network f is trained in an end-to-end manner to transform the deterministic input mat_i into a stochastic sentence representation. To capture multi-grained local context, filters with varying kernel sizes are used to extract semantic features which are aggregated through max-pooling operations. The architecture terminates in a dual-branch structure parameterized by θ_1 and θ_2.

These branches involve fully connected layers to simultaneously estimate the semantic feature vector μ_i and the representation uncertainty σ_i as defined in Eq. (1) and Eq. (2). To enable a differentiable operation, we apply the reparameterization trick to generate the latent vector $z_i = \mu_i + \epsilon \odot \sigma_i$, where ϵ follows a standard normal distribution. This stochastic vector z_i serves as the input for

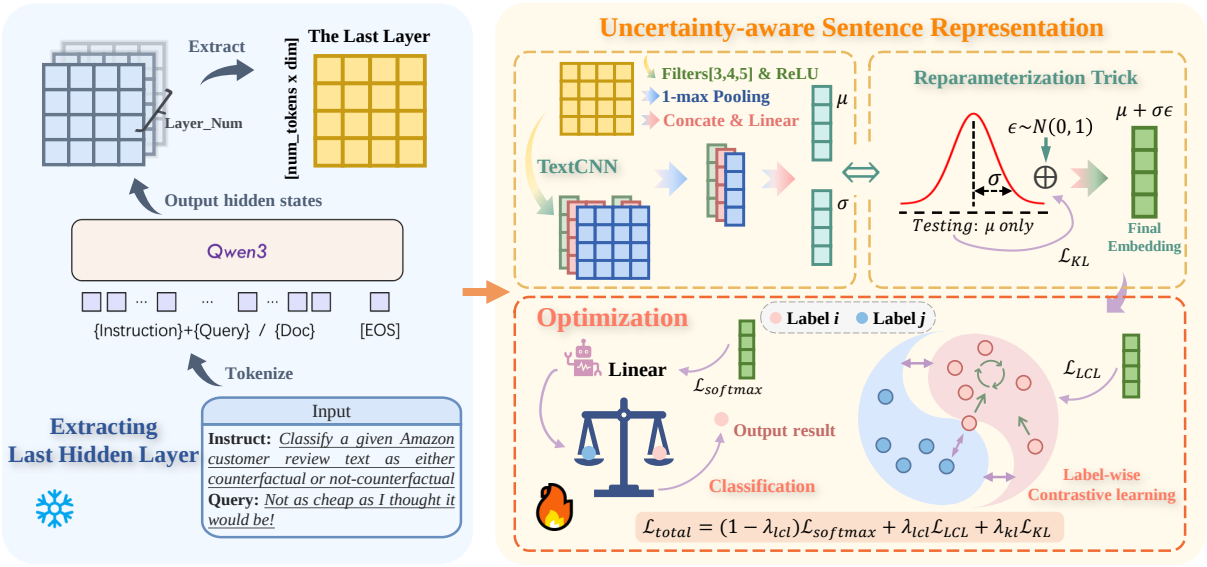

Figure 2: Framework of LLM-based uncertainty-aware contrastive sentence embedding.

the subsequent label-wise contrastive learning and classification modules.

$$\mu_i = f_{\theta_1}(mat_i) \tag{1}$$

$$\sigma_i = f_{\theta_2}(mat_i) \tag{2}$$

To further enhance the discriminability of the embeddings, we design a label-wise contrastive learning strategy, which operates directly on the stochastic representations generated at the reparameterization step. Specifically, each sampled vector z_i is normalized to yield $\hat{z}_i$. Unlike standard contrastive learning methods that rely on data augmentation, our approach exploits explicit class labels to construct positive pairs. For a given anchor sample x_i, let $\mathcal{P}(i)$ denote the set of indices for samples belonging to the same class within the batch. The objective function encourages the model to maximize the similarity between the anchor and its positive counterparts while simultaneously suppressing the similarity between the anchor and a sample from another class. The specific formulation is defined in Eq. (3), where τ serves as a temperature parameter to scale the dot products and n denotes the batch size.

$$\mathcal{L}_{LCL} = \frac{1}{n}\sum_{i=1}^{n} -\log \frac{\sum\limits_{k\in\mathcal{P}(i)} \exp(\hat{z}_i \cdot \hat{z}_k/\tau)}{\sum\limits_{j=1}^{n} \exp(\hat{z}_i \cdot \hat{z}_j/\tau)} \tag{3}$$

$$\mathcal{L}_{softmax} = \frac{1}{n}\sum_{i=1}^{n} -log \frac{exp(w_{y_i} \cdot z_i)}{\sum\limits_{c=1}^{k} exp(w_c \cdot z_i)} \tag{4}$$

$$\mathcal{L}_{KL} = KL[\mathcal{N}(z_i|\mu_i, \sigma_i^2) \parallel \mathcal{N}(\epsilon|0, \mathrm{I})] \tag{5}$$

$$\mathcal{L}_{total} = (1-\lambda_{lcl})\mathcal{L}_{softmax} + \lambda_{lcl}\mathcal{L}_{LCL} + \lambda_{kl}\mathcal{L}_{KL} \tag{6}$$

The model optimization is driven by a composite objective function, formulated as Eq. (6), which guides the simultaneous estimation of the feature representation μ_i and the uncertainty σ_i for the i-th sample. The objective function integrates three distinct components to strike a balance among classification accuracy, feature discriminability, and uncertainty-aware learning. Here, λ_{lcl} balances the classification loss with the label-wise contrastive learning term, while λ_{kl} weights the KL divergence regularization term for uncertainty estimation.

The first term $\mathcal{L}_{softmax}$ (Eq. (4)) maximizes the likelihood of predicting the correct class label y_i among k classes, serving as the primary supervision signal, where w_c represents the weight vector of the classifier for class c. Concurrently, the label-wise contrastive learning term $\mathcal{L}_{LCL}$ explicitly refines the feature space geometry. By enhancing intra-class compactness and inter-class separability, this term strengthens the separation between classes, making the classifier less sensitive to those ambiguous samples.

Crucially, the KL divergence term $\mathcal{L}_{KL}$ acts as a regulation by constraining the estimated distribution towards standard Gaussian distribution. This term serves a dual purpose. It both prevents the degeneration of stochastic representations into deterministic point estimates by penalizing negligible variances, and aligns the magnitude of predicted

Task (Abbr.)	Train	Test	Label
AmazonCounterfactual (Amz)	4018	670	2
Banking77 (Bank)	10003	3080	77
Imdb	25000	25000	2
MTOPDomain (MTOPD)	15667	4386	11
MassiveIntent (Mass-I)	11514	2974	60
MassiveScenario (Mass-S)	11514	2974	18
ToxicConversations (Toxic)	50000	50000	2
TweetSentimentExtraction (Tweet)	27481	3534	3

Table 1: Statistics and abbreviations of the selected MTEB(eng, v2) classification tasks

variance with the semantic ambiguity of the input sentence. Specifically, the model is learned to assign lower variance to semantically clear sentences and higher variance to ambiguous ones. This mechanism effectively makes the representation learning focus more on semantically clearer sentences by giving little corruption σ to their representations μ and to avoid overfitting those ambiguous sentences by corrupting the representations in larger degrees.

In the testing stage, the output of the pre-trained embedding model is taken as an input to our TextCNN that then outputs μ_i to serve as the final sentence representation of x_i for text classification. Overall, our proposed approach adapts a pre-trained model to a specific text classification task by integrating uncertainty-aware modeling with label-wise contrastive learning, thereby producing robust multi-grained local context representations.

5 Experiments

In this section, we evaluate the proposed approach experimentally in numerous text classification tasks, and compare it with the state-of-the-art methods in terms of the classification performance, and finally discuss the results.

5.1 Benchmarks and Tasks

To demonstrate the effectiveness of the feature representation produced by our method in text classification tasks, we conducted experiments based on LLMs. We evaluate our method using the English classification tasks from MTEB (English, v2) of the Massive Multilingual Text Embedding Benchmark (MMTEB) (Enevoldsen et al., 2025). MMTEB[2] is a large-scale, community-driven expansion of MTEB (Muennighoff et al., 2023), representing the largest multilingual collection of evaluation tasks for embedding models to date. Detailed statistics are presented in Table 1, where we list the abbreviations adopted in subsequent sections for brevity.

5.2 Baselines

We conduct a comprehensive comparison of our proposed approach against a diverse set of state-of-the-art text embedding models and commercial API services. Specifically, our evaluation includes prominent open-source baselines such as the Qwen3-Embedding series (Zhang et al., 2025a), embeddinggemma-300m (Vera et al., 2025), F2LLM-0.6B (Zhang et al., 2025b), and LGAI-Embedding-Preview (Choi et al., 2025). We also benchmark against three leading commercial APIs, namely, OpenAI's text-embedding-3-large, Google's Gemini-embedding (Lee et al., 2025b), and ByteDance's Seed1.5-Embedding.

5.3 Evaluation Setup

To demonstrate the universality of our method across different model architectures and pretraining paradigms, we conducted experiments on three distinct lightweight backbone models using an NVIDIA Tesla A100 GPU: Qwen3-Embedding-0.6B (denoted as Qwen3), F2LLM-0.6B (F2LLM), and embeddinggemma-300m (Gemma).

Consistent with the MTEB protocol, the pre-trained LLM backbones are frozen during training, input sequences along with their model-specific prompts[3] are truncated to a maximum length of 512 tokens. For the local context extractor, we employ a TextCNN architecture with kernel sizes of $\{3, 4, 5\}$ and a dropout rate of 0.5. The model is optimized using the AdamW optimizer for 3 epochs with a batch size of 64. We employ a linear learning rate warmup strategy. The learning rate is tuned specifically for each backbone (1e-4 for Qwen3, 8e-4 for others), regularization parameter λ_{LCL} is also tuned (0.6 for Qwen3, 0.7 and 0.9 for F2LLM and Gemma) and the temperature τ in contrastive loss is fixed at 2.0. A comprehensive description of the implementation details can be found in Appendix A.1.

In the inference stage, the reparameterization trick used in training is disabled. We utilize the estimated mean vector μ as the final sentence representation for classification, ensuring stable and reproducible predictions.

[2]The datasets can be obtained from https://huggingface.co/mteb/datasets

[3]Each model's prompt is publicly available on its HuggingFace homepage: https://huggingface.co/

Model	Amz	Bank	Imdb	MTOPD	Mass-I	Mass-S	Toxic	Tweet	Avg.
Leaderboard Baselines									
text-embedding-3-large	78.99	85.69	87.67	95.38	74.63	79.79	68.82	62.22	79.15
Gemini Embedding	92.69	94.27	94.98	99.27	88.46	92.08	88.75	69.88	90.05
Seed1.5-Embedding	91.40	91.70	96.98	99.28	87.81	93.26	86.65	71.95	89.88
LGAI-Embedding-Preview	93.18	91.38	96.81	98.25	82.40	85.41	92.76	79.58	89.97
Qwen3-Embedding-8B	93.94	87.27	97.37	98.39	85.87	89.98	91.65	78.98	90.43
Qwen3-Embedding-4B	93.73	86.34	97.16	97.78	85.02	88.80	91.44	78.45	89.84
Qwen3-Embedding-0.6B	91.46	81.01	95.44	95.96	80.43	83.59	82.13	76.05	85.76
F2LLM-0.6B	94.24	89.01	95.64	99.18	84.97	90.63	91.89	78.94	90.56
embeddinggemma-300m	90.07	91.45	92.92	99.11	85.79	91.54	82.93	66.59	87.55
Reproduced Baselines									
Qwen3-Embedding-0.6B	92.84	90.13	95.14	98.29	86.58	89.91	93.92	76.94	90.47
embeddinggemma-300m	92.84	91.72	92.28	99.13	87.30	91.90	94.06	73.29	90.31
F2LLM-0.6B	92.54	85.29	93.69	98.04	82.41	87.36	93.22	72.27	88.10
Our Method+Baselines									
Qwen3-Embedding-0.6B	93.88	93.47	95.64	99.20	89.14	92.23	94.43	78.38	92.05
F2LLM-0.6B	96.27	92.66	95.47	99.34	88.57	91.73	94.92	78.66	92.20
embeddinggemma-300m	95.37	93.86	94.36	99.29	89.71	92.54	94.69	77.96	**92.22**

Table 2: Performance of LLM-based models on eight classification tasks of the MTEB(eng, v2) benchmark.

5.4 Overall Performance

We evaluate the effectiveness of the proposed approach using classification accuracy across eight tasks from the MTEB benchmark. Table 2 summarizes the classification performance, where the scores of the standard baseline models are sourced from the online leaderboard[4]. The abbreviations of various datasets are denoted in Table 1. Furthermore, to ensure a fair comparison and isolate the improvements brought by our method, we also report the reproduced performance of several key baselines as detailed in Appendix A.1.

As indicated in Table 2, across all three backbone models, our method consistently and substantially improves performance over both their standard leaderboard scores and the reproduced baselines. Notably, the gains are observed on every individual task. While the base models often perform worse than larger competitors, our optimized lightweight variants not only close this performance gap but also surpass much larger models. Specifically, our method optimizing *embeddinggemma-300m* achieves a leading average accuracy of 92.22%, surpassing the Gemini Embedding and the much larger Qwen3-Embedding-8B. This effectiveness is further evidenced by the Qwen3-embedding series, where our method optimizing Qwen3-embedding-0.6B outperforms the standard Qwen3-embedding-4B and 8B models.

This consistent improvement underscores the generalizability and effectiveness of integrating uncertainty modeling with label-wise contrastive learning for sentence encoders. Specifically, the uncertainty-aware learning mechanism enables the model to be aware of ambiguous samples, while contrastive learning enhances inter-class separability. This synergy enhances representation robustness and avoids overfitting to ambiguous samples, effectively refining representations.

5.5 Ablation Studies

To investigate the individual contributions of our proposed modules and architectural choices, we conducted ablation studies on the Qwen3 backbone. Table 3 confirms the complementary benefits of Uncertainty-aware Context Learning (UCL) and Label-wise Contrastive Learning (LCL). Table 4 validates the superiority of TextCNN and LCL over alternative designs. Detailed experimental results among all three backbone models are given in A.3.

Impacts of Proposed Modules. The ablation results in Table 3 quantify the individual and combined contributions of the UCL and LCL modules across three strategies of extracting sentence embeddings, namely, EOS, Attention, and TextCNN. Adding either module individually leads to consistent performance improvements across all three sentence embedding extraction methods. When both UCL and LCL modules are combined, all backbones reach their best performance.

Specifically, LCL improves accuracy by explicitly optimizing intra-class compactness, while UCL enhances performance by regulating the learning

[4]Leaderboard results are retrieved from the online platform: https://huggingface.co/spaces/mteb/leaderboard

Components		Avg. Accuracy		
UCL	LCL	EOS	Attention	TextCNN
-	-	90.47	90.60	91.45
-	✓	91.20	91.25	91.77
✓	-	91.30	91.24	91.79
✓	✓	91.55	91.37	**92.05**

Table 3: Ablation analysis of the proposed modules (UCL and LCL) on different backbones.

Embedding Extraction		Contrastive Learning	
Method	Avg. Acc.	Method	Avg. Acc.
TextCNN	**91.45**	TextCNN (Base)	91.45
Mean	88.68	with SimCSE	91.50
MLP	89.74	with SupCon	91.44
Attention	90.60	with LCL	**91.77**

Table 4: Ablation studies on embedding extraction method and contrastive learning objectives.

process for ambiguous samples. Notably, TextCNN consistently achieves the highest scores in each configuration, benefiting more from both UCL and LCL modules. For TextCNN, the combination of UCL and LCL (i.e., our proposed method) achieves the best average accuracy of 92.05%, which highlights their complementary nature in enhancing the embedding space.

Effectiveness of Embedding Extraction. We validate the superiority of TextCNN-based local context embedding extraction over standard pooling method and attention-based aggregation. As shown in the left part of Table 4, TextCNN achieves average accuracy of 91.45%, substantially outperforming Mean Pooling (88.68%) and MLP (89.74%). Mean Pooling loses fine-grained positional information by simply averaging token embeddings, while the MLP-based approach, despite being trainable with the same number of parameters, fails to capture local n-gram patterns. More importantly, TextCNN also surpasses the Attention mechanism(90.60%), which assigns weights to token embeddings. The results confirm that explicitly capturing multi-grained local contexts yields richer semantic information than global aggregation methods. This representation enhancement is achieved via convolutional kernels of varying window sizes, which is particularly beneficial for sentences where local token interactions play a critical role.

Effect of Architectural Design on Performance Gains. To indicate that performance gains

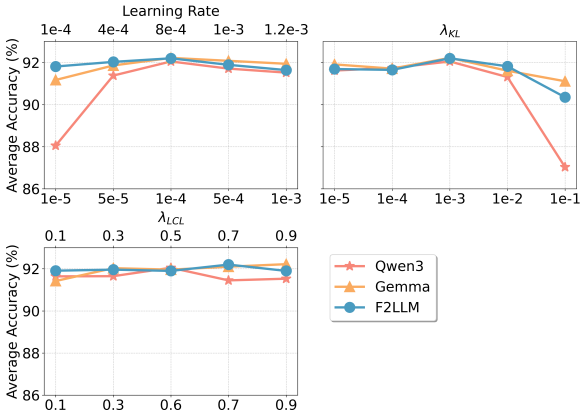

Figure 3: Sensitivity analysis of hyperparameters.

stem from our strategy for local context learning, we compared TextCNN with a non-linear MLP adapter having the exact same parameter count (≈ 19M). The MLP adapter achieves average accuracy of 89.74%, significantly lagging behind TextCNN. This result provides strong evidence that the effectiveness of our approach is derived from the specific architectural design capable of capturing local dependencies, rather than simply from the increase in the number of learnable parameters.

Comparison of Contrastive Learning Strategies. We further compare different contrastive learning strategies, as shown in the right part of Table 4. Our strategy of Label-wise Contrastive Learning achieves 91.77% average accuracy, compared to 91.50% for the unsupervised strategy (SimCSE) and 91.44% for the supervised strategy (SupCon) under the same baseline TextCNN encoder. While SimCSE relies solely on dropout-induced noise to construct positive pairs without exploiting label information, and SupCon explicitly utilizes class labels but processes all samples deterministically, LCL is designed to operate on stochastic representations derived from the uncertainty learning module. Our setting of uncertainty-aware contrastive learning allows the model to dynamically adjust decision boundaries based on semantic ambiguity, resulting in more robust class separability than traditional deterministic methods.

5.6 Impacts of Hyperparameters

The proposed objective function consists of a KL divergence component to regulate uncertainty estimation and a contrastive learning component to promote class discrimination. Consequently, the regularization parameter λ_{KL} and the contrastive weight λ_{LCL} are of particular importance, as they

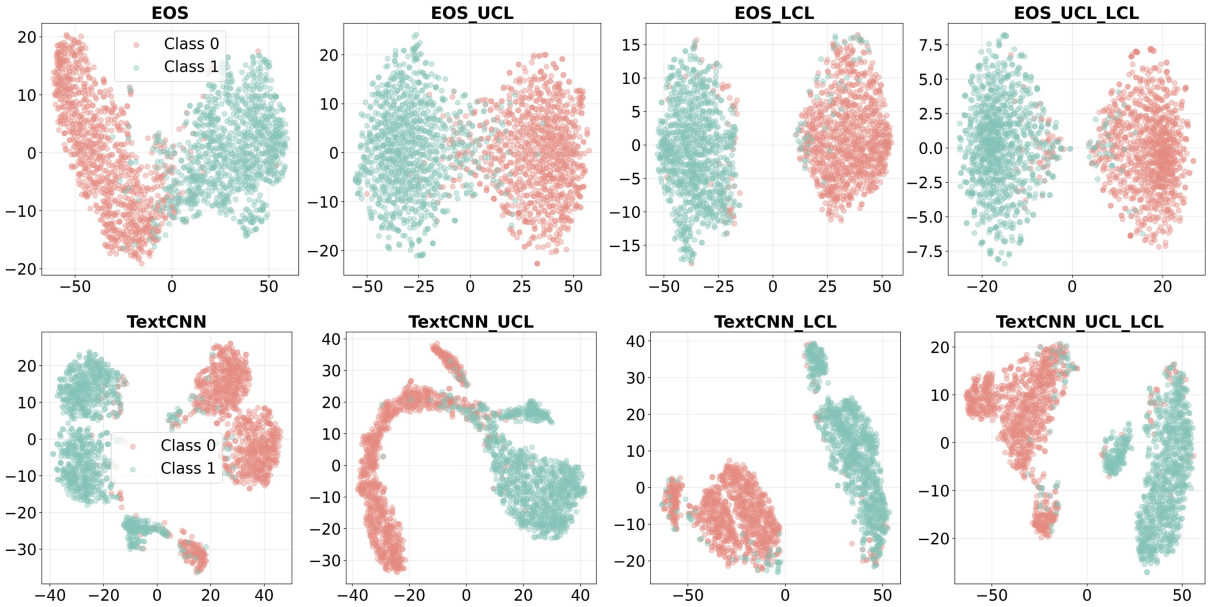

Figure 4: T-SNE visualization of representations learned by the Qwen3 model. The plot is generated using 2000 class-balanced samples randomly selected from the Imdb test set.

balance these auxiliary objectives with the primary classification loss. Given that the learning rate fundamentally determines the stability of optimization, we conduct sensitivity analysis of these three hyperparameters. Based on the results shown in Figure 3, which reports model performance under varying configurations, we establish the empirical foundation for our hyperparameter selection.

The optimal learning rate exhibits architecture dependent variation. Specifically, Qwen3 achieves its highest average accuracy at the learning rate of 1×10^{-4}, whereas Gemma and F2LLM perform optimally at 8×10^{-4}. Deviations from these values lead to non-trivial performance degradation. Regarding the KL regularization weight λ_{KL}, all three backbones consistently favor a value of 1×10^{-3}. Increasing λ_{KL} beyond this threshold results in a sharp decline in accuracy, with the most pronounced drop observed for Qwen3, whose performance falls to approximately 87.0% when λ_{KL} reaches 1×10^{-1}. In contrast, the method exhibits considerable robustness to the contrastive weight λ_{LCL} across the range of 0.3 to 0.9. The optimal λ_{LCL} values are 0.6 for Qwen3, 0.7 for F2LLM, and 0.9 for Gemma.

5.7 Analysis of uncertainty learning and semantic structure

To analyze the evolution of the feature space for the EOS and TextCNN architectures, the t-SNE visualizations are shown in Figure 4. The baseline model (EOS) results in dispersed sample distributions characterized by loose intra-class clustering and ambiguous boundaries. These observations suggest that deterministic models are hard to capture local semantics and sensitive to text ambiguity, resulting in insufficient intra-class compactness.

Incorporating uncertainty-aware context learning effectively enhances intra-class compactness and reduces inter-class overlaps. By identifying and regulating ambiguous samples with high uncertainty, the model achieves to make the confusion region involve mainly hard samples. Moreover, LCL alters the geometry of the representation space. The visualizations exhibit a clear expansion of inter-class margins, confirming that LCL helps enhance separability between classes.

The full method, integrating both uncertainty modeling and contrastive learning, achieves a synergistic refinement of the representation space. The final representations demonstrate both intra-class compactness and distinct inter-class separation. This representation enhancement is consistent across both EOS-based and TextCNN-based settings, validating the efficacy of our method in producing robust and discriminative features.

5.8 Analysis of uncertainty learning and prediction confidence

To evaluate the UCL module, we categorize testing samples into Easy, Semi-hard, and Hard levels based on the mean of the predicted uncertainty σ

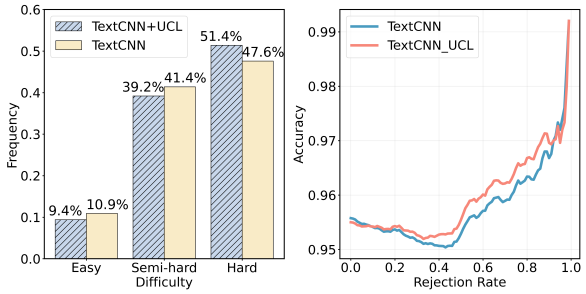

Figure 5: Analysis of uncertainty learning and prediction confidence using Qwen3 on the IMDB test set: (Left) Comparison of TextCNN predictions with and without UCL on challenging examples; (Right) Accuracy-Rejection Curve.

using thresholds of 0.8 and 1.0. The left of Figure 5 illustrates the distribution of misclassified samples. Compared to the TextCNN baseline, our method reduces the error proportion in the Easy and Semi-hard categories, resulting in a higher concentration of misclassifications within the Hard category. This shift indicates that the UCL module helps avoid overfitting by identifying and limiting the impact of ambiguous inputs. By paying less attention to hard samples, the decision boundary is optimized for samples with clear semantic structures.

Analysis of the reliability of uncertainty estimation via the Accuracy-Rejection Curve (Nadeem et al., 2009) is shown in the right part of Figure 5. We use uncertainty scores from TextCNN-UCL to guide sample rejection for both the baseline TextCNN and our proposed model. As a result, Figure 5 shows two upward trends where accuracy is improved with higher rejection rates, confirming that the learned uncertainty helps effectively detect low-confidence samples for the baseline.

5.9 Computational Efficiency

Metric	EOS	Ours
# Params	595.9 M	614.7 M
Latency (ms)	2.72	2.76
Throughput (samples/s)	367.0	362.1

Table 5: Efficiency comparison

To evaluate the practical deployability of the proposed framework, we analyze its computational efficiency, as presented in Table 5. Our proposed method introduces a parameter overhead of 3.16% compared to the baseline. Regarding inference speed, the inclusion of the UCL module leads to a marginal latency increase of 0.037ms, or 1.36% relative to the baseline, while still achieving a throughput exceeding 360 samples per second. These results demonstrate that the framework achieves substantial accuracy improvements with negligible computational cost, confirming its viability for latency-sensitive real-time scenarios.

6 Conclusion

In this paper, we have proposed an approach of uncertainty-aware contrastive sentence embedding with local context representation, which involves adapting pre-trained models to text classification tasks, through simultaneously learning to transform a word embedding matrix into a contextualized sentence vector and to quantify the representation uncertainty of the sentence. In particular, we have designed an end-to-end strategy to learn multi-grained local context representations in the setting of TextCNN, where the representation learning is done by taking our uncertainty-aware strategy to address the ambiguity existing in some sentences and adopting our contrastive learning strategy to enhance intra-class compactness and inter-class separability. The experimental results show that our approach outperforms those state-of-the-art embedding methods in terms of text classification accuracy, through incorporating uncertainty-aware contrastive sentence embedding with local context representation into a pre-trained model. The results also indicate that our approach can help smaller embedding models outperform larger models.

Limitations

Our current work focuses on processing of English language, but some other languages may need tokenization in specific ways that are very dissimilar to the ones used for English text, such as Chinese and Japanese. From this point of view, the proposed approach may not be directly applicable to some other languages, so it is worthy of future studies on how to adapt the proposed approach to various tokenization ways used for different languages. Moreover, our objective function is designed particularly to drive uncertainty-aware contrastive sentence embedding for text classification tasks but may not be adaptive to language generation tasks. Therefore, in the future, it is worth to explore how to design objective functions that are adaptive to text generation in the setting of uncertainty-aware learning.

Ethics Statement

Our work focuses on sentence embedding for text classification tasks, where all the data sets used for our experiments are public. We have cited and used the data sets in the ways that are consistent with the intended ones specified in the licenses. Moreover, this research does not involve human subjects or animals. Therefore, we do not anticipate any ethical issues arising from the work presented in this paper.

Acknowledgments

We acknowledge the support of the Shandong Provincial Natural Science Foundation (Grant ZR2025QC1591), the National Natural Science Foundation of China (Grant 62576221), the Guangdong Provincial Natural Science Foundation (2025A1515010288), the Guangdong Provincial Key Laboratory (Grant 2023B1212060076) and the Shenzhen Science and Technology Program (Grant ZDSYS20220527171400002).

References

Moloud Abdar, Farhad Pourpanah, Sadiq Hussain, Dana Rezazadegan, Li Liu, Mohammad Ghavamzadeh, Paul Fieguth, Xiaochun Cao, Abbas Khosravi, U. Rajendra Acharya, Vladimir Makarenkov, and Saeid Nahavandi. 2021. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information Fusion*, 76:243–297.

Parishad BehnamGhader, Vaibhav Adlakha, Marius Mosbach, Dzmitry Bahdanau, Nicolas Chapados, and Siva Reddy. 2024. LLM2Vec: Large language models are secretly powerful text encoders. *arXiv preprint arXiv:2404.05961*.

Jie Chang, Zhonghao Lan, Changmao Cheng, and Yichen Wei. 2020. Data uncertainty learning in face recognition. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5710–5719, Online. IEEE.

Chanyeol Choi, Junseong Kim, Seolhwa Lee, Jihoon Kwon, Sangmo Gu, Yejin Kim, Minkyung Cho, and Jy-yong Sohn. 2024. Linq-embed-mistral technical report. *arXiv preprint arXiv:2412.03223*.

Jooyoung Choi, Hyun Kim, Hansol Jang, Changwook Jun, Kyunghoon Bae, Hyewon Choi, Stanley Jungkyu Choi, Honglak Lee, and Chulmin Yun. 2025. Lgai-embedding-preview technical report. *CoRR*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 4171–4186, New Orleans, Louisiana. Association for Computational Linguistics.

Kenneth Enevoldsen, Isaac Chung, Imene Kerboua, Márton Kardos, Ashwin Mathur, David Stap, Jay Gala, Wissam Siblini, Dominik Krzeminski, Genta Indra Winata, and 1 others. 2025. MMTEB: Massive multilingual text embedding benchmark. In *International Conference on Learning Representations*.

Nicholas Frosst, Nicolas Papernot, and Geoffrey Hinton. 2019. Analyzing and improving representations with the soft nearest neighbor loss. In *International Conference on Machine Learning*, pages 2012–2020. PMLR.

Tianyu Gao, Xingcheng Yao, and Danqi Chen. 2021. SimCSE: Simple contrastive learning of sentence embeddings. In *2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021*, pages 6894–6910, Punta Cana, Dominican Republic. Association for Computational Linguistics (ACL).

Anchun Gui and Han Xiao. 2024. Multi-level multilingual semantic alignment for zero-shot cross-lingual transfer learning. *Neural Networks*, 173:106217.

Dong Guo, Faming Wu, Feida Zhu, Fuxing Leng, Guang Shi, Haobin Chen, Haoqi Fan, Jian Wang, Jianyu Jiang, Jiawei Wang, and 1 others. 2025. Seed1. 5-vl technical report. *arXiv preprint arXiv:2505.07062*.

Saumya Gupta, Yikai Zhang, Xiaoling Hu, Prateek Prasanna, and Chao Chen. 2024. Topology-aware uncertainty for image segmentation. *Advances in Neural Information Processing Systems*, 36.

Chao Huang, Yushu Shi, Bob Zhang, and Ke Lyu. 2024. Uncertainty-aware prototypical learning for anomaly detection in medical images. *Neural Networks*, 175:106284.

Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. 2020. Supervised contrastive learning. *Advances in Neural Information Processing Systems*, 33:18661–18673.

Diederik P. Kingma and Max Welling. 2014. Auto-encoding variational bayes. In *2nd International Conference on Learning Representations (ICLR)*, page 14, Banff, Canada.

Puneet Kumar and Balasubramanian Raman. 2022. A bert based dual-channel explainable text emotion recognition system. *Neural Networks*, 150:392–407.

Chankyu Lee, Rajarshi Roy, Mengyao Xu, Jonathan Raiman, Mohammad Shoeybi, Bryan Catanzaro, and Wei Ping. 2025a. NV-Embed: Improved techniques for training LLMs as generalist embedding models. In *The Thirteenth International Conference on Learning Representations*.

Jinhyuk Lee, Feiyang Chen, Sahil Dua, Daniel Cer, Madhuri Shanbhogue, Iftekhar Naim, Gustavo Hernández Ábrego, Zhe Li, Kaifeng Chen, Henrique Schechter Vera, and 1 others. 2025b. Gemini embedding: Generalizable embeddings from gemini. *arXiv preprint arXiv:2503.07891*.

Wanhua Li, Xiaoke Huang, Jiwen Lu, Jianjiang Feng, and Jie Zhou. 2021. Learning probabilistic ordinal embeddings for uncertainty-aware regression. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13896–13905, Nashville, USA. IEEE.

Zehan Li, Xin Zhang, Yanzhao Zhang, Dingkun Long, Pengjun Xie, and Meishan Zhang. 2023. Towards general text embeddings with multi-stage contrastive learning. *arXiv preprint arXiv:2308.03281*.

Tao Liu, Jiahao Liu, Dong Li, and Shan Tan. 2025a. Bayesian deep-learning structured illumination microscopy enables reliable super-resolution imaging with uncertainty quantification. *Nature Communications*, 16(1):5027.

Xiaolu Liu, Ruizi Yang, Song Wang, Wentong Li, Junbo Chen, and Jianke Zhu. 2025b. Uncertainty-instructed structure injection for generalizable hd map construction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 22359–22368.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. RoBERTa: A robustly optimized BERT pretraining approach. *arXiv preprint arXiv:1907.11692*.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Proceedings of the 27th Conference on Neural Information Processing Systems*, pages 3111–3119, Lake Tahoe, USA. Curran Associates.

Niklas Muennighoff, Hongjin Su, Liang Wang, Nan Yang, Furu Wei, Tao Yu, Amanpreet Singh, and Douwe Kiela. 2025. Generative representational instruction tuning. In *The Thirteenth International Conference on Learning Representations*.

Niklas Muennighoff, Nouamane Tazi, Loic Magne, and Nils Reimers. 2023. MTEB: Massive text embedding benchmark. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics.

Malik Sajjad Ahmed Nadeem, Jean-Daniel Zucker, and Blaise Hanczar. 2009. Accuracy-rejection curves (ARCs) for comparing classification methods with a reject option. In *Machine Learning in Systems Biology*, pages 65–81. PMLR.

Jianmo Ni, Gustavo Hernandez Abrego, Noah Constant, Ji Ma, Keith Hall, Daniel Cer, and Yinfei Yang. 2022. Sentence-T5: Scalable sentence encoders from pretrained text-to-text models. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 1864–1874, Dublin, Ireland. Association for Computational Linguistics.

Zhijie Nie, Zhangchi Feng, Mingxin Li, Cunwang Zhang, Yanzhao Zhang, Dingkun Long, and Richong Zhang. 2024. When text embedding meets large language model: a comprehensive survey. *arXiv preprint arXiv:2412.09165*.

Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. GloVe: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pages 1532–1543, Doha, Qatar. Association for Computational Linguistics.

Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. 2018. Improving language understanding by generative pre-training. Technical Report 4, OpenAI, San Francisco, CA, USA.

Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3982–3992, Hong Kong, China. Association for Computational Linguistics (ACL).

Haojing Shen, Sihong Chen, Ran Wang, and Xizhao Wang. 2023a. Adversarial learning with cost-sensitive classes. *IEEE Transactions on Cybernetics*, 53(8):4855–4866.

Lingfeng Shen, Haiyun Jiang, Lemao Liu, and Shuming Shi. 2023b. Sen2Pro: A probabilistic perspective to sentence embedding from pre-trained language model. In *Proceedings of the 8th Workshop on Representation Learning for NLP (RepL4NLP 2023)*, pages 315–333.

Dianmo Sheng, Dongdong Chen, Zhentao Tan, Qiankun Liu, Qi Chu, Tao Gong, Bin Liu, Jing Han, Wenbin Tu, Shengwei Xu, and Nenghai Yu. 2025. Unicl-sam: Uncertainty-driven in-context segmentation with part prototype discovery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 20201–20211.

Jacob Mitchell Springer, Suhas Kotha, Daniel Fried, Graham Neubig, and Aditi Raghunathan. 2024. Repetition improves language model embeddings. *arXiv preprint arXiv:2402.15449*.

Henrique Schechter Vera, Sahil Dua, Biao Zhang, Daniel Salz, Ryan Mullins, Sindhu Raghuram Panyam, Sara Smoot, Iftekhar Naim, Joe Zou, Feiyang Chen, and 1 others. 2025. Embeddinggemma: Powerful and lightweight text representations. *arXiv preprint arXiv:2509.20354*.

Ranwan Wu, Tian-Zhu Xiang, Guo-Sen Xie, Rongrong Gao, Xiangbo Shu, Fang Zhao, and Ling Shao. 2025. Uncertainty-aware transformer for referring camouflaged object detection. *IEEE Transactions on Image Processing*, 34:5341–5354.

Shohei Yoda, Hayato Tsukagoshi, Ryohei Sasano, and Koichi Takeda. 2024. Sentence representations via gaussian embedding. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 418–425.

Wenqiao Zhang and Zheqi Lv. 2024. Revisiting the domain shift and sample uncertainty in multi-source active domain transfer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16751–16761.

Xin Zhang, Yanzhao Zhang, Dingkun Long, Wen Xie, Ziqi Dai, Jialong Tang, Huan Lin, Baosong Yang, Pengjun Xie, Fei Huang, Meishan Zhang, Wenjie Li, and Min Zhang. 2024. mGTE: Generalized long-context text representation and reranking models for multilingual text retrieval. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: Industry Track*, pages 1393–1412, Miami, Florida, US. Association for Computational Linguistics.

Yanzhao Zhang, Mingxin Li, Dingkun Long, Xin Zhang, Huan Lin, Baosong Yang, Pengjun Xie, An Yang, Dayiheng Liu, Junyang Lin, and 1 others. 2025a. Qwen3 embedding: Advancing text embedding and reranking through foundation models. *arXiv preprint arXiv:2506.05176*.

Ziyin Zhang, Zihan Liao, Hang Yu, Peng Di, and Rui Wang. 2025b. F2LLM technical report: Matching SOTA embedding performance with 6 million open-source data. *arXiv preprint arXiv:2510.02294*.

Xinping Zhao, Xinshuo Hu, Zifei Shan, Shouzheng Huang, Yao Zhou, Xin Zhang, Zetian Sun, Zhenyu Liu, Dongfang Li, Xinyuan Wei, and 1 others. 2025. KaLm-Embedding-V2: Superior training techniques and data inspire a versatile embedding model. *arXiv preprint arXiv:2506.20923*.

A Appendix

A.1 Implementation Details

In our experimental setup, we adopt a parameter-efficient fine-tuning strategy to adapt large language models for sentence embedding tasks. Specifically, the parameters of the pre-trained backbones are frozen to preserve their general linguistic capabilities. Optimization is exclusively applied to the downstream components, including the local context extractor and the final classifier.

To ensure a rigorous and fair comparison, we align our evaluation protocol with the standard MTEB evaluation process, treating the baseline setup as a supervised linear probe on frozen embeddings. Specifically, for the EOS-based baselines, we first extract the embedding corresponding to the EOS token from each input sequence without updating it during training. We then pass this static EOS token embedding through a trainable linear layer, using the resulting projection for classification. This process is analogous to training a logistic regression classifier on top of static embeddings, as done in MTEB, ensuring that both the baseline and our method benefit from the same level of task-specific supervision. The evaluation procedure is shown in Figure 6.

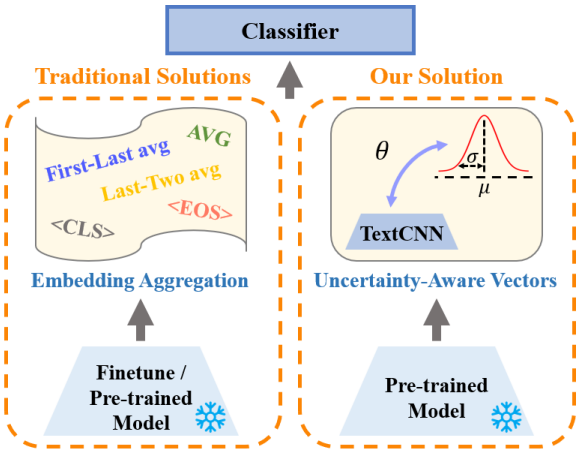

Figure 6: Compare the testing process of our solution with that of traditional solutions.

We standardize the architecture for all EOS-based ablation settings (including the standard baseline and variants integrated with auxiliary objectives). For the Qwen3 and F2LLM models, we extract the EOS token embedding (or the CLS token embedding for Gemma) and project it through a trainable linear transformation layer. Since the backbone is frozen, this projection head serves as the critical learnable component that maps the static global representation into an adaptive feature space. This design is essential as it provides the necessary trainable parameters for the label-wise contrastive learning objective to effectively optimize the embedding geometry. In contrast, our proposed method is designed by replacing this linear projection with a TextCNN-based local context extractor. This module processes the sequence of the last hidden states, mapping them into a multivariate Gaussian distribution defined by a mean vector μ and a variance vector σ.

In the testing stage, the stochastic sampling

Model Name	Parameters	Dim.
Commercial API Services		
text-embedding-3-large	–	3072
Gemini-embedding	–	768
Seed1.5-Embedding	–	1024
Open-Source Baselines		
Qwen3-Embedding-0.6B	0.6B	1024
Qwen3-Embedding-4B	4B	2560
Qwen3-Embedding-8B	8B	4096
embeddinggemma-300m	0.3B	768
F2LLM-0.6B	0.6B	1024
LGAI-Embedding-Preview	–	4096
Proposed Methods (Ours)		
Ours (Qwen3-0.6B)	0.6B + 19.0M	1024
Ours (Gemma-300m)	0.3B + 10.6M	768
Ours (F2LLM-0.6B)	0.6B + 19.0M	1024

Table 6: Overview of the baseline models and commercial APIs used in our evaluation. 'Dim.' denotes the dimension of the output sentence embedding. For our proposed methods, the parameter count is explicitly denoted as 'Backbone (Frozen) + Added Modules (Trainable)', highlighting the parameter-efficient nature of our approach.

mechanism is disabled to ensure deterministic outputs. Consequently, the estimated mean vector μ, which captures the rich semantic content of the sentence, serves as the input for classification.

A.2 Model Parameters Analysis of Baselines

We conduct a comprehensive comparison of our proposed approach against a diverse set of state-of-the-art text embedding models and commercial API services. Specifically, our evaluation includes prominent open-source baselines such as the Qwen3-Embedding series (Zhang et al., 2025a), embeddinggemma-300m (Vera et al., 2025), F2LLM-0.6B (Zhang et al., 2025b), and LGAI-Embedding-Preview (Choi et al., 2025). We also benchmark against three leading commercial APIs, namely, OpenAI's text-embedding-3-large, Google's Gemini-embedding (Lee et al., 2025b), and ByteDance's Seed1.5-Embedding.

A critical aspect of our experimental implementation is the adherence to a parameter-efficient fine-tuning strategy. As shown in the bottom section of Table 6, we implement our proposed framework on top of three lightweight backbones, namely, Qwen3-Embedding-0.6B, embeddinggemma-300m, and F2LLM-0.6B. During training, we freeze the parameters of the pretrained LLM backbones to preserve their generalization capabilities. Optimization is exclusively applied to the proposed uncertainty-aware TextCNN extractor and the classifier in our full setting. Consequently, our method introduces a negligible number of additional trainable parameters—approximately 19.0M for Qwen3 and F2LLM and 10.6M for Gemma. This design enables the extraction of rich local contexts and uncertainty modeling, resulting in a parameter overhead of approximately 3% relative to the base encoders.

A.3 Ablation Studies

In order to evaluate the individual contributions and generalizability of the proposed modules, we conducted ablation studies among all three backbones: Qwen3-Embedding-0.6B, embeddinggemma-300m, and F2LLM-0.6B. The following tables present the more detailed ablation results corresponding to the main text. Results are summarized in Tables 7, 8, 9, and 10.

We first quantify the contributions of UCL and LCL across three embedding extraction methods (Table 7). Adding either module individually improves performance, and combining both achieves the best results. Notably, TextCNN consistently achieves the highest scores, with the combination of UCL and LCL reaching 92.05% average accuracy, highlighting their complementary benefits.

We validate the superiority of TextCNN over pooling and attention methods (Table 8). TextCNN achieves 91.45% average accuracy, substantially outperforming Mean Pooling (88.68%) and MLP (89.74%). More importantly, TextCNN also surpasses Attention (90.60%). The aforementioned results confirm that capturing multi-grained local contexts yields richer semantic information than global aggregation methods.

Effect of Architectural Design and Contrastive Objectives. To validate the effectiveness of our architectural design, we compared TextCNN with an MLP adapter of identical parameter count. The MLP adapter achieves only 89.74% accuracy, significantly lower than TextCNN (91.45%), confirming that performance gains stem from architectural design rather than parameter count. Furthermore, under the same TextCNN setup, our Label-wise Contrastive Learning (LCL) achieves 91.77% average accuracy, outperforming SimCSE (91.50%) and SupCon (91.44%). Unlike deterministic contrastive methods, LCL operates on stochastic representations from the uncertainty learning module, dynamically adjusting decision boundaries based on semantic ambiguity for better separability.

Generalization to Other Backbones. The pro-

Dataset	EOS				Attention				TextCNN			
	Base	+LCL	+UCL	+Both	Base	+LCL	+UCL	+Both	Base	+LCL	+UCL	+Both
AmazonCounterfactual	92.84	94.03	93.28	94.03	91.79	94.22	93.75	94.22	92.54	93.73	93.53	93.88
Banking77	90.13	92.11	91.59	91.95	91.40	91.80	91.80	91.44	92.66	93.15	93.21	93.47
Imdb	95.14	95.38	95.49	95.45	94.85	94.88	95.35	95.39	95.49	95.58	95.53	95.64
MTOPDomain	98.29	98.77	98.86	99.00	98.88	98.68	98.81	98.75	99.04	99.02	99.03	99.20
MassiveIntent	86.58	87.22	88.10	88.70	87.79	88.01	88.08	87.84	88.35	88.84	88.80	89.14
MassiveScenario	89.91	90.72	91.32	91.39	90.35	91.00	90.35	91.13	91.39	91.83	91.67	92.23
ToxicConversations	93.92	93.62	93.93	93.94	94.43	94.28	93.51	94.33	94.11	94.36	94.43	94.43
TweetSentiment	76.94	77.73	77.79	77.96	75.30	77.13	78.27	77.84	78.01	77.65	78.14	78.38
Average	90.47	91.20	91.30	91.55	90.60	91.25	91.24	91.37	91.45	91.77	91.79	**92.05**

Table 7: Ablation study on **Qwen3-Embedding-0.6B**. Classification accuracy (%) on eight datasets. We group variants by architecture: EOS-based, Attention-base and TextCNN-based modules.The variants include the baseline and its enhancements with Label-wise Contrastive Learning (LCL) and Uncertainty-aware Context Learning (UCL).

Dataset	Embedding Extraction				Contrastive Learning			
	Mean	MLP	Attention	TextCNN	Base	+SimCSE	+SupCon	+LCL
AmazonCounterfactual	93.58	90.30	91.79	92.54	92.54	92.34	91.94	93.73
Banking77	84.58	91.21	91.40	92.66	92.66	92.71	92.69	93.15
Imdb	93.53	94.42	94.85	95.49	95.49	95.33	95.62	95.58
MTOPDomain	97.63	98.61	98.88	99.04	99.04	99.20	99.02	99.02
MassiveIntent	83.93	84.75	87.79	88.35	88.35	89.50	88.33	88.84
MassiveScenario	89.48	88.89	90.35	91.39	91.39	91.63	91.27	91.83
ToxicConversations	93.19	94.04	94.43	94.11	94.11	93.55	94.38	94.36
TweetSentiment	73.49	75.68	75.30	78.01	78.01	77.76	78.30	77.65
Average	88.68	89.74	90.60	**91.45**	91.45	91.50	91.44	**91.77**

Table 8: Detailed experimental results of all reported classification tasks evaluating embedding extraction methods and contrastive learning objectives on **Qwen3-Embedding-0.6B**.

Dataset	Base	CLS	TextCNN			
	(LeaderBoard)	(Reproduct)	Base	+LCL	+UCL	+Full
AmazonCounterfactual	90.07	92.84	91.94	95.22	94.78	**95.37**
Banking77	91.45	91.72	92.69	93.34	93.48	**93.86**
Imdb	92.92	92.28	94.00	94.28	94.09	**94.36**
MTOPDomain	99.11	99.13	99.20	99.25	99.24	**99.29**
MassiveIntent	85.79	87.30	88.20	88.33	89.44	**89.71**
MassiveScenario	91.54	91.90	91.96	92.43	92.29	**92.54**
ToxicConversations	82.93	94.06	94.51	93.95	94.58	**94.69**
TweetSentimentExtraction	66.59	73.29	76.49	77.90	76.90	**77.96**

Table 9: Ablation study on **embeddinggemma-300m** for reporting the accuracy (%) across different modules.

Dataset	Base	EOS	TextCNN			
	(LeaderBoard)	(Reproduct)	Base	+LCL	+UCL	+Full
AmazonCounterfactual	94.24	92.54	95.52	95.67	95.67	**96.27**
Banking77	89.01	85.29	91.92	91.30	91.94	**92.66**
Imdb	95.64	93.69	95.23	95.33	95.36	**95.47**
MTOPDomain	99.18	98.04	99.09	99.18	99.22	**99.34**
MassiveIntent	84.97	82.41	87.06	88.03	87.93	**88.57**
MassiveScenario	90.63	87.36	91.13	91.39	91.43	**91.73**
ToxicConversations	91.89	93.22	94.83	94.63	94.78	**94.92**
TweetSentimentExtraction	78.94	72.27	78.24	78.64	78.61	**78.66**

Table 10: Ablation study on **F2LLM-0.6B** for reporting the accuracy (%) across different module combinations.

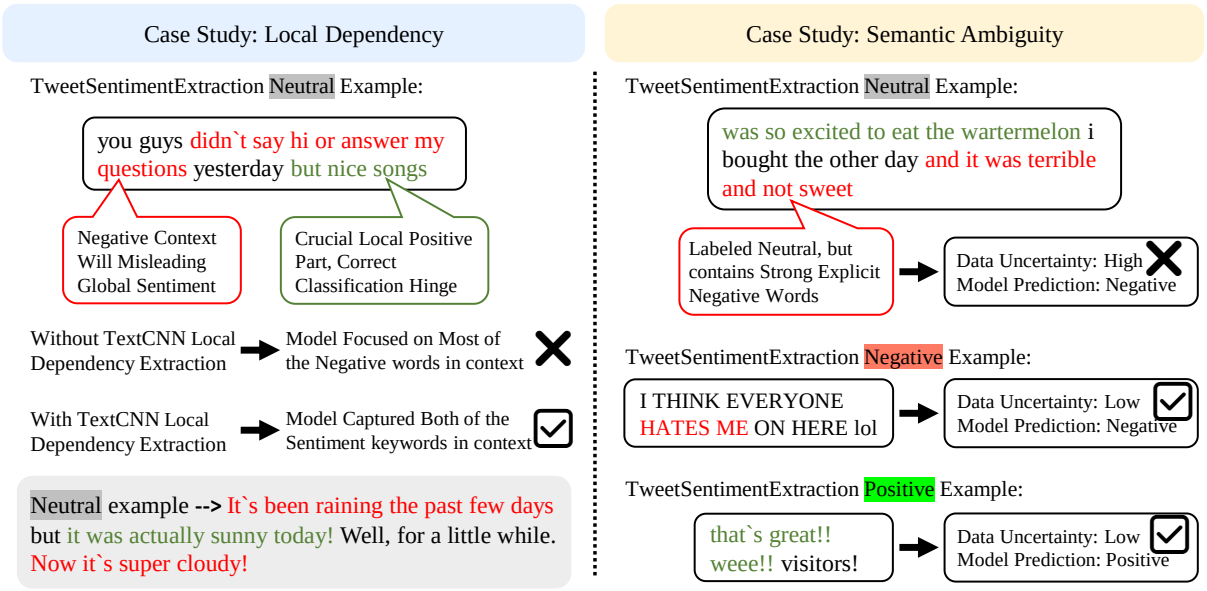

Figure 7: Case Study: Illustrating Local Dependency and Semantic Ambiguity.

posed modules also demonstrate consistent effectiveness on embeddinggemma-300m (Table 9) and F2LLM-0.6B (Table 10). For embeddinggemma-300m, the full model (TextCNN + LCL + UCL) achieves the highest accuracy on 7 out of 8 datasets, with an average improvement of +2.16% over the EOS variant termed as Base (Reproduct). Similarly, on F2LLM-0.6B, the full model outperforms both the leaderboard Base and the reimplemented EOS baseline on all datasets, achieving the best overall accuracy (e.g., 96.27% on AmazonCounterfactual and 99.34% on MTOPDomain). These results confirm that the proposed method is not specific to Qwen3-Embedding-0.6B but generalizes well across different backbone architectures.

A.4 Case Studies

We sampled several cases from the TweetSentimentExtraction test set as shown in Figure 7. To illustrate the motivation of extracting local dependency, consider the positive sample: "you guys didn't say hi or answer my questions yesterday but nice songs." The sentence begins with a negative complaint but abruptly flips with the local compliment "but nice songs" on which making a correct classification strongly depends. Without local dependency extraction, a model is easily misled by the initial complaint, leading to an incorrect result.

A neutral example relying on local chunks (contexts) is: "It's been raining the past few days but it was actually sunny today! Well, for a little while. Now it's super cloudy!". This sentence rapidly shifts among "raining" (negative), "sunny" (positive), and "cloudy" (negative). A global representation may not capture these conflicting signals, losing crucial details. In contrast, local context extraction captures these distinct semantic chunks to identify the overall sentiment. To establish a baseline, our module assigns low variance to sentences with highly certain meanings. For example, the explicitly positive sentence "that's great!! weee!! visitors!" and the strongly negative sentence "I THINK EVERYONE HATES ME ON HERE lol" both express clear and unambiguous emotions. Our model accurately assigns them low uncertainty, confirming that it recognizes clear semantic signals.

In contrast, our module assigns high variance to texts with semantic ambiguity. Consider the sample: "was so excited to eat the wartermelon i bought the other day and it was terrible and not sweet". This sentence involves explicit negative words ("terrible" and "not sweet") to express disappointment. Its sentiment looks negative to some extent, but it is labeled as neutral in the dataset. By assigning a high variance to this sample, our uncertainty learning module prevents model from overfitting the hard sample.