Novelty Assessment Report

Paper: Let LLMs Speak Embedding Languages: Generative Text Embeddings via Iterative Contrastive Refinement
PDF URL: https://openreview.net/pdf?id=okjogxO1Fu
Venue: ICLR 2026 Conference Submission
Year: 2026
Report Generated: 2026-01-01

Abstract

Existing large language model (LLM)-based embeddings typically adopt an encoder-only paradigm, treating LLMs as static feature extractors and overlooking their core gener- ative strengths. We introduce GIRCSE (Generative Iterative Refinement for Contrastive Sentence Embeddings), a novel framework that leverages autoregressive generation to iter- atively refine semantic representations. By producing sequences of soft tokens optimized under a contrastive objective, GIRCSE captures latent concepts and implicit semantics that encoder-only methods often miss. To guide this process, we propose an Iterative Contrastive Refinement (ICR) objective that encourages each refinement step to yield bet- ter representations. Extensive experiments show that GIRCSE outperforms strong LLM- based embedding baselines on the MTEB embedding benchmark. Moreover, GIRCSE ex- hibits an emergent test-time scaling property: generating more tokens at inference steadily improves embedding quality. Our results establish generative iterative refinement as a new paradigm for representation learning.

Core Task Landscape

This paper addresses: **Generative Text Embeddings via Iterative Contrastive Refinement**

A total of **36 papers** were analyzed and organized into a taxonomy with **17 categories**.

Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Generative Refinement for Text Representation Learning**
- **Contrastive Learning for Static Embeddings**
- **Iterative Refinement for Downstream Generation Tasks**
- **Clustering and Classification with Contrastive and Iterative Methods**
- **Specialized Applications and Cross-Domain Methods**

Complete Taxonomy Tree

- Generative Text Embeddings via Iterative Contrastive Refinement Survey Taxonomy
- Generative Refinement for Text Representation Learning
  - LLM-Based Generative Embedding Frameworks ★ (3 papers)
  - [0] Let LLMs Speak Embedding Languages: Generative Text Embeddings via Iterative Contrastive Refinement (Anon et al., 2026) View paper
  - [13] Resource-Efficient Adaptation of Large Language Models for Text Embeddings via Prompt Engineering and Contrastive Fine-tuning (Benedikt Roth, 2025) View paper
  - [23] Large Language Models can Contrastively Refine their Generation for Better Sentence Representation Learning (Huiming Wang, 2024) View paper
  - Iterative Process Refinement for Agent and Task Learning (3 papers)
  - [3] Watch Every Step! LLM Agent Learning via Iterative Step-level Process Refinement (Li Cheng, 2024) View paper
  - [10] Eipe-text: Evaluation-guided iterative plan extraction for long-form narrative text generation (You Wang, 2023) View paper
  - [26] Reading, writing, and refining for text summarization (Shichao, 2024) View paper
  - Diffusion and Probabilistic Generative Models for Retrieval (1 papers)
  - [5] Diffatr: Diffusion-based generative modeling for audio-text retrieval (Xin Yifei, 2024) View paper
- Contrastive Learning for Static Embeddings
  - General-Purpose Text Embedding Models (3 papers)
  - [22] Sentence Embeddings in NLI with Iterative Refinement Encoders (Aarne Talman, 2022) View paper
  - [32] Conan-embedding: General Text Embedding with More and Better Negative Samples (Li Shiyu, 2024) View paper
  - [33] Towards General Text Embeddings with Multi-stage Contrastive Learning (Li, 2023) View paper
  - Embedding Refinement via Post-hoc Contrastive Learning (1 papers)
  - [28] Can Contrastive Learning Refine Embeddings (Liu Lihui, 2024) View paper
  - Hierarchical and Multi-Granularity Contrastive Learning (2 papers)
  - [11] Keywords and instances: A hierarchical contrastive learning framework unifying hybrid granularities for text generation (Li, 2022) View paper
  - [34] Long: Keywords and Instances: A Hierarchical Contrastive Learning Framework Unifying Hybrid Granularities for Text Generation (Chang Jin-xiong, 2022) View paper
  - Multimodal Contrastive Embedding Learning (4 papers)
  - [8] ClusterE-ZSL: a novel cluster-based embeddings for enhanced zero-shot learning in contrastive pre-training cross-model retrieval (Umair Tariq, 2024) View paper

    - [16] HAMLET-FFD: Hierarchical Adaptive Multi-modal Learning Embeddings Transformation for Face Forgery Detection (Jialei Cui, 2025) View paper
    - [20] Progressive Local Alignment for Medical Multimodal Pre-training (Yan Huimin, 2025) View paper
    - [27] Ensuring Pre-Fusion Modality Consistency: A New Approach to Multimodal Sentiment Detection (Yulou Shu, 2025) View paper
- Iterative Refinement for Downstream Generation Tasks
    - Contrastive Decoding and Generation Quality Enhancement (2 papers)
    - [2] Multi-amateur contrastive decoding for text generation (Jaydip Sen, 2025) View paper
    - [9] Iter-ahmcl: Alleviate hallucination for large language model via iterative model-level contrastive learning (Wu Huiwen, 2024) View paper
    - Multimodal Report and Description Generation (3 papers)
    - [12] Look, Imitate and Refine: A Hierarchical Multimodel Retrieval Augmented Vision-Language Model for Radiology Report Generation (Kai Chen, 2024) View paper
    - [24] Bootstrapping Large Language Models for Radiology Report Generation (Chang Liu, 2024) View paper
    - [25] Bootstrapping Large Language Models for Radiology Report Generation | VIDEO (Chen Wei-Dong, 2024) View paper
    - Retrieval-Augmented and Feedback-Driven Generation (2 papers)
    - [6] An iterative feedback mechanism for improving natural language class descriptions in open-vocabulary object detection (Karker, 2025) View paper
    - [19] MERLIN: Multimodal Embedding Refinement via LLM-based Iterative Navigation for Text-Video Retrieval-Rerank Pipeline (Donghoon Han, 2024) View paper
    - 3D and 4D Content Generation with Alignment (1 papers)
    - [29] PLA4D: Pixel-Level Alignments for Text-to-4D Gaussian Splatting (Miao, 2024) View paper
- Clustering and Classification with Contrastive and Iterative Methods
    - Unsupervised Text Clustering with Iterative Refinement (2 papers)
    - [14] Cone: Unsupervised Contrastive Opinion Extraction (Zhao, 2023) View paper
    - [36] CEIL: A General Classification-Enhanced Iterative Learning Framework for Text Clustering (Zhao Mingjun, 2023) View paper
    - Intent Discovery and Classification via Contrastive Learning (4 papers)
    - [4] Clear Up Confusion: Iterative Differential Generation for Fine-grained Intent Detection with Contrastive Feedback (F Zhang, 2025) View paper
    - [7] Small sample-based adaptive text classification through iterative and contrastive description refinement (Avadhanam Udayaadithya, 2025) View paper
    - [30] Controllable Discovery of Intents: Incremental Deep Clustering Using Semi-Supervised Contrastive Learning (Rawat, 2023) View paper
    - [35] Induce Spoken Dialog Intents via Deep Unsupervised Context Contrastive Clustering (Ting-Wei Wu, 2022) View paper
    - Domain-Specific Contrastive Clustering (3 papers)
    - [1] Refinement Contrastive Learning of Cell-Gene Associations for Unsupervised Cell Type Identification (Liang Peng, 2025) View paper
    - [17] Learning Audio Foundation Models for Reasoning (Deshmukh, 2025) View paper
    - [18] Image Clustering through Invariant Representation Learning and Text Criteria Conditioning (è¶□ì□,í□□, 2025) View paper
- Specialized Applications and Cross-Domain Methods
    - Embedding Inversion and Privacy Analysis (1 papers)
    - [21] Text Embeddings Reveal (Almost) As Much As Text (Morris, 2023) View paper
    - Graph Neural Networks with LLM-Enhanced Iterative Refinement (1 papers)
    - [15] Are LLMs Better GNN Helpers? Rethinking Robust Graph Learning under Deficiencies with Iterative Refinement (Wang Zhaoyan, 2025) View paper
    - Weakly-Supervised Segmentation with Text Guidance (1 papers)
    - [31] Weakly-Supervised Text Instance Segmentation (Xinyan Zu, 2023) View paper

Narrative

Core task: generative text embeddings via iterative contrastive refinement. The field structure reflects a broad interest in combining generative modeling, contrastive objectives, and iterative improvement to produce high-quality text representations. The taxonomy organizes work into several main branches. Generative Refinement for Text Representation Learning explores how large language models and generative frameworks can produce or refine embeddings through multiple passes, often leveraging prompt-based strategies or feedback loops. Contrastive Learning for Static Embeddings focuses on learning discriminative representations by contrasting positive and negative pairs, typically in a single-stage manner. Iterative Refinement for Downstream Generation Tasks examines multi-step processes that improve outputs like summaries or plans by repeatedly revising intermediate results. Clustering and Classification with Contrastive and Iterative Methods addresses unsupervised or semi-supervised scenarios where iterative updates and contrastive signals help discover latent structure or refine class boundaries. Finally, Specialized Applications and Cross-Domain Methods capture domain-specific adaptations, such as medical imaging or audio-text retrieval, where iterative contrastive techniques are tailored to unique data modalities.

A particularly active line of work involves LLM-based generative embedding frameworks that marry the expressive power of large models with contrastive objectives, as seen in Generative Text Embeddings[0] and Prompt Contrastive Embeddings[13]. These approaches contrast with more traditional static contrastive methods like Contrastive Sentence Representation[23], which refine embeddings without explicit generative iteration. Another vibrant area is iterative refinement for generation tasks, where methods such as Step-level Process Refinement[3] and Iterative Description Feedback[6] progressively enhance outputs through feedback loops, raising questions about how many refinement cycles are optimal and how to balance computational cost with quality gains. Generative Text Embeddings[0] sits squarely within the generative refinement branch, emphasizing iterative contrastive updates to produce embeddings that capture nuanced semantic distinctions. Compared to Prompt Contrastive Embeddings[13], which also leverages prompts for contrastive learning, and Contrastive Sentence Representation[23], which focuses on static sentence-level contrasts, Generative Text Embeddings[0] distinguishes itself by integrating multiple refinement passes to iteratively sharpen representation quality, positioning it at the intersection of generative modeling and contrastive learning.

Related Works in Same Category

The following **2 sibling papers** share the same taxonomy leaf node with the original paper:

1. Resource-Efficient Adaptation of Large Language Models for Text Embeddings via Prompt Engineering and Contrastive Fine-tuning

Authors: Benedikt Roth, Qiu Tianming, Stephan Rappensperger, Tianming Qiu, WÃ¶rmann, et al. (10 authors total) | **Year/Venue**: 2025 • arXiv.org | **URL**: View paper

Abstract

Large Language Models (LLMs) have become a cornerstone in Natural Language Processing (NLP), achieving impressive performance in text generation. Their token-level representations capture rich, human-aligned semantics. However, pooling these vectors into a text embedding discards crucial information. Nevertheless, many non-generative downstream tasks, such as clustering, classification, or retrieval, still depend on accurate and controllable sentence- or document-level embeddings. We explore sev...

Relationship Analysis

Both papers belong to the LLM-Based Generative Embedding Frameworks category, exploring how to leverage large language models' capabilities for text embeddings. While the original paper (GIRCSE) introduces an end-to-end generative framework that autoregressively generates soft tokens optimized via iterative contrastive refinement, the candidate paper focuses on resource-efficient adaptation strategies including prompt engineering, token aggregation techniques, and contrastive fine-tuning on synthetically generated pairs. The key difference is that GIRCSE proposes a novel generative iterative refinement mechanism with test-time scaling properties, whereas the candidate paper explores more conventional adaptation methods (pooling strategies and prompt engineering) for decoder-only LLMs without the iterative generation component.

---

2. Large Language Models can Contrastively Refine their Generation for Better Sentence Representation Learning

Authors: Huiming Wang, Zhaodonghui Li, Liying Cheng, De Wen Soh, Lidong Bing | **Year/Venue**: 2024 | **URL**: View paper

Abstract

Recently, large language models (LLMs) have emerged as a groundbreaking technology and their unparalleled text generation capabilities have sparked interest in their application to the fundamental sentence representation learning task. Existing methods have explored utilizing LLMs as data annotators to generate synthesized data for training contrastive learning based sentence embedding models such as SimCSE. However, since contrastive learning models are sensitive to the quality of sentence pair...

Relationship Analysis

Both papers belong to the LLM-Based Generative Embedding Frameworks category, leveraging large language models' generative capabilities for text embeddings. While the original paper (GIRCSE) uses autoregressive generation to iteratively refine embeddings through soft token generation with contrastive objectives, the candidate paper (MultiCSR) uses LLMs as data generators to create synthesized sentence pairs for training separate contrastive learning models like SimCSE. The key difference is that GIRCSE performs end-to-end generative embedding within the LLM itself, whereas MultiCSR uses LLMs to generate training data for traditional encoder-based contrastive models.

Contributions Analysis

Overall novelty summary. The paper introduces GIRCSE, a framework that leverages autoregressive generation to iteratively refine sentence embeddings under a contrastive objective. It resides in the 'LLM-Based Generative Embedding Frameworks' leaf, which contains only three papers total, indicating a relatively sparse and emerging research direction. This leaf sits within the broader 'Generative Refinement for Text Representation Learning' branch, distinguishing itself from the more populated 'Contrastive Learning for Static Embeddings' category by emphasizing generative processes during embedding formation rather than single-pass encoding.

The taxonomy reveals that neighboring work includes 'Iterative Process Refinement for Agent and Task Learning' (three papers) and 'Diffusion and Probabilistic Generative Models for Retrieval' (one paper), both exploring iterative or generative mechanisms but for different purposes—agent trajectories and cross-modal retrieval, respectively. The sibling papers in the same leaf focus on prompt-based contrastive embeddings and generative text embeddings, suggesting that GIRCSE shares conceptual ground with these approaches but diverges by explicitly modeling iterative refinement steps. The broader 'Contrastive Learning for Static Embeddings' branch (thirteen papers across four leaves) represents a more mature, crowded area focused on discriminative objectives without generative iteration.

Among thirty candidates examined, none were found to clearly refute any of the three core contributions: the GIRCSE framework, the Iterative Contrastive Refinement objective, and the test-time scaling property. Each contribution was assessed against ten candidates, with zero refutable overlaps identified. This suggests that within the limited search scope, the specific combination of generative iteration, contrastive refinement, and emergent test-time scaling appears relatively unexplored. However, the small candidate pool and the sparse taxonomy leaf indicate that the field is still nascent, making it difficult to draw definitive conclusions about novelty without broader literature coverage.

Based on the top-thirty semantic matches and the sparse taxonomy structure, the work appears to occupy a relatively underexplored niche at the intersection of generative modeling and contrastive embedding learning. The absence of refutable prior work within this limited scope is encouraging, but the small number of sibling papers and the emerging nature of the research direction suggest that the field is still consolidating. A more exhaustive search or future work may reveal additional connections as this area matures.

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: GIRCSE framework for generative text embeddings

Description: The authors propose GIRCSE, a framework that uses autoregressive generation to produce sequences of soft tokens optimized under contrastive objectives, enabling iterative refinement of text embeddings rather than single-pass encoding. This approach captures latent concepts and implicit semantics that encoder-only methods often miss.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Keywords and instances: A hierarchical contrastive learning framework unifying hybrid granularities for text generation

URL: View paper

Brief Assessment

Hierarchical Contrastive Granularities[11] focuses on text generation tasks (paraphrasing, dialogue, storytelling) using keyword-level and instance-level contrasts, not on iterative autoregressive refinement of sentence embeddings for representation learning.

---

2. Enhancing scientific literature summarization via contrastive learning and chain-of-thought prompting

URL: View paper

Brief Assessment

Contrastive Chain-of-Thought Summarization[58] focuses on scientific literature summarization using chain-of-thought prompting for abstractive text generation, not on autoregressive generation for iterative refinement of contrastive sentence embeddings as in GIRCSE.

---

3. TEACH: A Contrastive Knowledge Adaptive Distillation Framework for Classical Chinese Understanding

URL: View paper

Brief Assessment

Classical Chinese Distillation[63] focuses on knowledge distillation for classical Chinese understanding tasks (word sense disambiguation and translation), not on generative text embeddings or contrastive sentence representation learning. The candidate employs contrastive learning for distilling translation capabilities, whereas the original develops autoregressive generation for iterative embedding refinement.

---

4. A Study on Improving Japanese Writing Skills by Constructing Japanese Syntactic Analysis and Generation Technology Using Computational Methods and â□¦

URL: View paper

Brief Assessment

Japanese Syntactic Generation[64] focuses on Japanese syntactic analysis and text generation for improving writing skills, not on autoregressive generation for iterative refinement of contrastive sentence embeddings. The candidate addresses dependency parsing and text embellishment in Japanese language education, which is a completely different domain and technical approach from the original paper's generative text embedding framework.

---

5. Multilingual pre-training model-assisted contrastive learning neural machine translation

URL: View paper

Brief Assessment

Multilingual Contrastive Translation[62] focuses on neural machine translation for low-resource ethnic languages using contrastive learning to address exposure bias, not on generative text embeddings or iterative refinement of sentence representations.

---

6. Context Matters: Enhancing Sequential Recommendation with Context-aware Diffusion-based Contrastive Learning

URL: View paper

Brief Assessment

Context-aware Diffusion Contrastive[57] focuses on sequential recommendation systems using diffusion models for data augmentation in contrastive learning, not on generative text embeddings or autoregressive refinement of sentence representations.

---

7. Large Language Models can Contrastively Refine their Generation for Better Sentence Representation Learning

URL: View paper

Brief Assessment

Contrastive Sentence Representation[23] focuses on using LLMs to generate synthetic training data for contrastive learning models, not on autoregressive generation for iterative embedding refinement. The candidate uses LLMs as data annotators to create sentence pairs for training separate base models, while GIRCSE directly generates soft tokens within the embedding model itself for iterative refinement.

---

8. Protip: Progressive tool retrieval improves planning

URL: View paper

Brief Assessment

Progressive Tool Retrieval[60] focuses on tool retrieval for multi-step planning tasks using contrastive learning for task decomposition, not on generative text embeddings via autoregressive refinement of sentence representations.

---

9. Sequential contrastive learning for progressive knowledge tracing

URL: View paper

Brief Assessment

Sequential Contrastive Tracing[61] focuses on knowledge tracing in educational contexts, not on generative text embeddings or autoregressive refinement of sentence representations. The candidate paper's domain and methodology are fundamentally different from the original paper's contribution.

---

10. Aligning semantic in brain and language: A curriculum contrastive method for electroencephalography-to-text generation

URL: View paper

Brief Assessment

Curriculum Contrastive EEG[59] focuses on EEG-to-text generation using contrastive learning to align brain signals with semantic representations, not on autoregressive generation for iterative refinement of text embeddings from language models.

Contribution 2: Iterative Contrastive Refinement (ICR) objective

Description: The authors introduce ICR, a training objective that provides contrastive supervision at every generation step and enforces progressive embedding quality improvement. This objective guides the generative embedding process toward high-quality representations through stepwise contrastive loss and iterative refinement regularization.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Incremental model enhancement via memory-based contrastive learning

URL: View paper

Brief Assessment

Memory-based Contrastive Enhancement[43] focuses on incremental model enhancement from sequentially available training splits, which is a different problem setting from the ORIGINAL paper's iterative generation-based embedding refinement with stepwise contrastive supervision.

---

2. Supervised contrastive replay: Revisiting the nearest class mean classifier in online class-incremental continual learning

URL: View paper

Brief Assessment

Supervised Contrastive Replay[37] focuses on continual learning with supervised contrastive loss for memory replay, not on iterative generative refinement of embeddings through autoregressive token generation.

---

3. Efficient Event Camera Data Pretraining with Adaptive Prompt Fusion

URL: View paper

Brief Assessment

Adaptive Prompt Fusion[45] focuses on event camera data pretraining through spatiotemporal information fusion prompting, not on contrastive objectives with stepwise supervision for progressive embedding quality improvement in text embeddings.

---

4. PETFormer-SCL: a supervised contrastive learning-guided CNNâ□□transformer hybrid network for Parkinsonism classification from FDG-PET

URL: View paper

Brief Assessment

Supervised Contrastive Parkinsonism[39] focuses on medical image classification for Parkinsonism using contrastive learning in a CNN-Transformer hybrid architecture, not on iterative generative embedding refinement for text representations.

---

5. OCL: Ordinal Contrastive Learning for Imputating Features with Progressive Labels

URL: View paper

Brief Assessment

Ordinal Contrastive Imputation[40] focuses on medical imaging feature imputation for Alzheimer's disease progression, using ordinal contrastive loss to align samples by disease stage. This differs fundamentally from ICR's stepwise contrastive supervision for iterative text embedding refinement.

---

6. U-MARVEL: Unveiling Key Factors for Universal Multimodal Retrieval via Embedding Learning with MLLMs

URL: View paper

Brief Assessment

Universal Multimodal Retrieval[46] focuses on multimodal retrieval using MLLMs with standard contrastive learning, not on iterative generation with stepwise supervision for progressive embedding refinement.

---

7. Semi-supervised feature contrast incremental learning framework for bearing fault diagnosis with limited labeled samples

URL: View paper

Brief Assessment

Semi-supervised Incremental Diagnosis[41] focuses on fault diagnosis with feature contrastive loss for semi-supervised learning, not on iterative refinement objectives for progressive embedding quality improvement in generative text embeddings.

---

8. Dual contrastive learning framework for incremental text classification

URL: View paper

Brief Assessment

Dual Contrastive Incremental[42] focuses on incremental text classification with dual contrastive learning (task-agnostic and task-specific), not on generative embeddings with stepwise supervision for progressive refinement during autoregressive generation.

---

9. Progressive negative enhancing contrastive learning for image dehazing and beyond

URL: View paper

Brief Assessment

Progressive Negative Dehazing[44] focuses on image dehazing with progressive negative sample updates for contrastive learning in computer vision, not on iterative embedding refinement for text representations in NLP.

---

10. Diffmm: Multi-modal diffusion model for recommendation

URL: View paper

Brief Assessment

Multi-modal Diffusion Recommendation[38] focuses on multi-modal diffusion models for recommendation systems with cross-modal contrastive learning, not on iterative contrastive refinement objectives for progressive embedding quality improvement in text representation learning.

Contribution 3: Test-time scaling property for text embeddings

Description: The authors demonstrate that GIRCSE shows consistent embedding quality improvements with increased refinement steps at inference time, representing a novel scaling paradigm for embedding models analogous to test-time compute scaling in reasoning LLMs. This allows controllable performance gains through adjustable generation length.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. CAT-TPT: Class-Agnostic Text-based Test-time Prompt Tuning for Vision-Language Models

URL: View paper

Brief Assessment

Class-Agnostic Test-time Prompting[56] focuses on vision-language models and test-time prompt tuning for visual classification tasks, not text embedding quality improvements through iterative generation steps.

---

2. Scaling up test-time compute with latent reasoning: A recurrent depth approach

URL: View paper

Brief Assessment

Recurrent Depth Reasoning[50] focuses on test-time scaling for reasoning tasks through recurrent latent computation in language models, not text embeddings. The original paper's contribution is specific to embedding quality improvement through iterative refinement of semantic representations.

---

3. Testnuc: Enhancing test-time computing approaches through neighboring unlabeled data consistency

URL: View paper

Brief Assessment

Neighboring Unlabeled Consistency[53] focuses on test-time computing for classification tasks using neighboring unlabeled data consistency, not on test-time scaling for embedding quality through iterative refinement steps as in GIRCSE.

---

4. A Survey on Test-Time Scaling in Large Language Models: What, How, Where, and How Well?

URL: View paper

Brief Assessment

Test-Time Scaling Survey[51] focuses on test-time scaling in LLMs for reasoning tasks (mathematics, coding, open-ended Q&A), not on text embedding models or representation learning through iterative refinement.

---

5. SANA 1.5: Efficient Scaling of Training-Time and Inference-Time Compute in Linear Diffusion Transformer

URL: View paper

Brief Assessment

SANA Linear Diffusion[48] focuses on inference-time scaling for text-to-image generation models through repeated sampling strategies, not text embedding models or semantic representation quality improvements.

---

6. Scaling embedding layers in language models

URL: View paper

Brief Assessment

Scaling Embedding Layers[55] focuses on scaling input embedding layers through n-gram embeddings with off-accelerator memory storage, not on test-time scaling through iterative refinement steps at inference. The scaling mechanisms are fundamentally different.

---

7. Efficiently scaling transformer inference

URL: View paper

Brief Assessment

Efficient Transformer Inference[47] focuses on scaling transformer inference through partitioning strategies and hardware optimization for large language models, not on test-time scaling for embedding quality through iterative refinement.

---

8. Testnuc: Enhancing test-time computing approaches and scaling through neighboring unlabeled data consistency

URL: View paper

Brief Assessment

Neighboring Unlabeled Consistency[52] focuses on test-time scaling through leveraging neighboring unlabeled data for classification tasks, not on iterative refinement of text embeddings through increased generation steps as in GIRCSE.

---

9. Inference scaling for long-context retrieval augmented generation

URL: View paper

Brief Assessment

Long-context Retrieval Scaling[54] focuses on scaling inference computation for retrieval augmented generation (RAG) in long-context LLMs, not on test-time scaling for text embedding models through iterative refinement steps.

---

10. Investigating test-time scaling with reranking for machine translation

URL: View paper

Brief Assessment

Test-time Reranking Translation[49] focuses on machine translation quality improvement through best-of-n candidate selection, not text embedding refinement. The scaling mechanism (generating multiple translation candidates vs. iterative embedding refinement) and application domain (MT vs. embeddings) are fundamentally different.

Appendix: Text Similarity Detection

No high-similarity text segments were detected across any compared papers.

References

- [0] Let LLMs Speak Embedding Languages: Generative Text Embeddings via Iterative Contrastive Refinement View paper
- [1] Refinement Contrastive Learning of Cell-Gene Associations for Unsupervised Cell Type Identification View paper
- [2] Multi-amateur contrastive decoding for text generation View paper
- [3] Watch Every Step! LLM Agent Learning via Iterative Step-level Process Refinement View paper
- [4] Clear Up Confusion: Iterative Differential Generation for Fine-grained Intent Detection with Contrastive Feedback View paper
- [5] Diffatr: Diffusion-based generative modeling for audio-text retrieval View paper
- [6] An iterative feedback mechanism for improving natural language class descriptions in open-vocabulary object detection View paper
- [7] Small sample-based adaptive text classification through iterative and contrastive description refinement View paper
- [8] ClusterE-ZSL: a novel cluster-based embeddings for enhanced zero-shot learning in contrastive pre-training cross-model retrieval View paper
- [9] Iter-ahmcl: Alleviate hallucination for large language model via iterative model-level contrastive learning View paper
- [10] Eipe-text: Evaluation-guided iterative plan extraction for long-form narrative text generation View paper
- [11] Keywords and instances: A hierarchical contrastive learning framework unifying hybrid granularities for text generation View paper
- [12] Look, Imitate and Refine: A Hierarchical Multimodel Retrieval Augmented Vision-Language Model for Radiology Report Generation View paper
- [13] Resource-Efficient Adaptation of Large Language Models for Text Embeddings via Prompt Engineering and Contrastive Fine-tuning View paper

- [14] Cone: Unsupervised Contrastive Opinion Extraction View paper
- [15] Are LLMs Better GNN Helpers? Rethinking Robust Graph Learning under Deficiencies with Iterative Refinement View paper
- [16] HAMLET-FFD: Hierarchical Adaptive Multi-modal Learning Embeddings Transformation for Face Forgery Detection View paper
- [17] Learning Audio Foundation Models for Reasoning View paper
- [18] Image Clustering through Invariant Representation Learning and Text Criteria Conditioning View paper
- [19] MERLIN: Multimodal Embedding Refinement via LLM-based Iterative Navigation for Text-Video Retrieval-Rerank Pipeline View paper
- [20] Progressive Local Alignment for Medical Multimodal Pre-training View paper
- [21] Text Embeddings Reveal (Almost) As Much As Text View paper
- [22] Sentence Embeddings in NLI with Iterative Refinement Encoders View paper
- [23] Large Language Models can Contrastively Refine their Generation for Better Sentence Representation Learning View paper
- [24] Bootstrapping Large Language Models for Radiology Report Generation View paper
- [25] Bootstrapping Large Language Models for Radiology Report Generation | VIDEO View paper
- [26] Reading, writing, and refining for text summarization View paper
- [27] Ensuring Pre-Fusion Modality Consistency: A New Approach to Multimodal Sentiment Detection View paper
- [28] Can Contrastive Learning Refine Embeddings View paper
- [29] PLA4D: Pixel-Level Alignments for Text-to-4D Gaussian Splatting View paper
- [30] Controllable Discovery of Intents: Incremental Deep Clustering Using Semi-Supervised Contrastive Learning View paper
- [31] Weakly-Supervised Text Instance Segmentation View paper
- [32] Conan-embedding: General Text Embedding with More and Better Negative Samples View paper
- [33] Towards General Text Embeddings with Multi-stage Contrastive Learning View paper
- [34] Long: Keywords and Instances: A Hierarchical Contrastive Learning Framework Unifying Hybrid Granularities for Text Generation View paper
- [35] Induce Spoken Dialog Intents via Deep Unsupervised Context Contrastive Clustering View paper
- [36] CEIL: A General Classification-Enhanced Iterative Learning Framework for Text Clustering View paper
- [37] Supervised contrastive replay: Revisiting the nearest class mean classifier in online class-incremental continual learning View paper
- [38] Diffmm: Multi-modal diffusion model for recommendation View paper
- [39] PETFormer-SCL: a supervised contrastive learning-guided CNNâtransformer hybrid network for Parkinsonism classification from FDG-PET View paper
- [40] OCL: Ordinal Contrastive Learning for Imputating Features with Progressive Labels View paper
- [41] Semi-supervised feature contrast incremental learning framework for bearing fault diagnosis with limited labeled samples View paper
- [42] Dual contrastive learning framework for incremental text classification View paper
- [43] Incremental model enhancement via memory-based contrastive learning View paper
- [44] Progressive negative enhancing contrastive learning for image dehazing and beyond View paper
- [45] Efficient Event Camera Data Pretraining with Adaptive Prompt Fusion View paper
- [46] U-MARVEL: Unveiling Key Factors for Universal Multimodal Retrieval via Embedding Learning with MLLMs View paper
- [47] Efficiently scaling transformer inference View paper
- [48] SANA 1.5: Efficient Scaling of Training-Time and Inference-Time Compute in Linear Diffusion Transformer View paper
- [49] Investigating test-time scaling with reranking for machine translation View paper
- [50] Scaling up test-time compute with latent reasoning: A recurrent depth approach View paper
- [51] A Survey on Test-Time Scaling in Large Language Models: What, How, Where, and How Well? View paper
- [52] Testnuc: Enhancing test-time computing approaches and scaling through neighboring unlabeled data consistency View paper
- [53] Testnuc: Enhancing test-time computing approaches through neighboring unlabeled data consistency View paper
- [54] Inference scaling for long-context retrieval augmented generation View paper
- [55] Scaling embedding layers in language models View paper
- [56] CAT-TPT: Class-Agnostic Text-based Test-time Prompt Tuning for Vision-Language Models View paper
- [57] Context Matters: Enhancing Sequential Recommendation with Context-aware Diffusion-based Contrastive Learning View paper
- [58] Enhancing scientific literature summarization via contrastive learning and chain-of-thought prompting View paper
- [59] Aligning semantic in brain and language: A curriculum contrastive method for electroencephalography-to-text generation View paper
- [60] Protip: Progressive tool retrieval improves planning View paper
- [61] Sequential contrastive learning for progressive knowledge tracing View paper
- [62] Multilingual pre-training model-assisted contrastive learning neural machine translation View paper
- [63] TEACH: A Contrastive Knowledge Adaptive Distillation Framework for Classical Chinese Understanding View paper
- [64] A Study on Improving Japanese Writing Skills by Constructing Japanese Syntactic Analysis and Generation Technology Using Computational Methods and â¦ View paper