Novelty Assessment Report

Paper: Energy-Regularized Sequential Model Editing on Hyperspheres
PDF URL: https://openreview.net/pdf?id=CHsdtzCip6
Venue: ICLR 2026 Conference Submission
Year: 2026
Report Generated: 2025-12-30

Abstract

Large language models (LLMs) require constant updates to remain aligned with evolving real-world knowledge. Model editing offers a lightweight alternative to retraining, but sequential editing that updates the LLM knowledge through multiple successive edits often destabilizes representations and induces catastrophic forgetting. In this work, we seek to better understand and mitigate performance degradation caused by sequential editing. We hypothesize that hyperspherical uniformity, a property that maintains uniform distribution of neuron weights on a hypersphere, helps the model remain stable, retain prior knowledge, while still accommodate new updates. We use Hyperspherical Energy (HE) to quantify neuron uniformity during editing, and examine its correlation with editing performance. Empirical studies across widely used editing methods reveals a strong correlation between HE dynamics and editing performance, with editing failures consistently coinciding with uncontrolled HE fluctuations. We further theoretically prove that HE dynamics impose a lower bound on the degradation of pretrained knowledge, highlighting why HE stability is crucial for knowledge retention. Motivated by these insights, we propose SPHERE (Sparse Projection for Hyperspherical Energy-Regularized Editing), an HE-driven regularization strategy that stabilizes neuron weight distributions, ultimately preserving prior knowledge while enabling reliable sequential updates. Specifically, SPHERE identifies a sparse space complementary to the principal hyperspherical directions of the pretrained weight matrices and projects new knowledge onto it, attenuating perturbations on the principal directions. Extensive experiments on LLaMA3 (8B) and Qwen2.5 (7B) show that SPHERE outperforms the best baseline in editing capability by an average of 16.41%, while most faithfully preserving general model performance, thereby offering a principled path toward reliable large-scale knowledge editing.

Core Task Landscape

This paper addresses: **Sequential Model Editing for Large Language Models**

A total of **50 papers** were analyzed and organized into a taxonomy with **27 categories**.

Taxonomy Overview

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

- **Sequential Editing Methods and Architectures**
- **Evaluation and Analysis of Sequential Editing**
- **Contextual and Retrieval-Based Knowledge Update**
- **Specialized Editing Applications and Extensions**
- **Related Knowledge Update and Reasoning Paradigms**
- **Cross-Domain and Multimodal Applications**
- **Foundational Techniques and Surveys**

Complete Taxonomy Tree

- Sequential Model Editing for Large Language Models Survey Taxonomy
- Sequential Editing Methods and Architectures
  - Parameter-Modifying Sequential Editing
  - Neuron-Level and Layer-Targeted Editing (2 papers)
    - [2] Explainable and efficient editing for large language models (Tianyu Zhang, 2025) View paper
    - [28] Neuron-Level Sequential Editing for Large Language Models (Jiang HouCheng, 2024) View paper
  - Orthogonal Subspace and Projection-Based Editing ★ (2 papers)
    - [0] Energy-Regularized Sequential Model Editing on Hyperspheres (Anon et al., 2026) View paper
    - [17] O-edit: Orthogonal subspace editing for language model sequential editing (Cai Yuchen, 2024) View paper
  - Model Merging for Knowledge Integration (1 papers)
    - [16] Model Merging for Knowledge Editing (Wu Xian, 2025) View paper
  - Parameter-Preserving Sequential Editing
  - Adapter-Based Knowledge Injection (1 papers)
    - [12] Selective Knowledge Injection via Adapter Modules in Largeâ□Scale Language Models (Hongye Zheng, 2025) View paper
  - Dual-Memory Architectures (1 papers)
    - [3] Wise: Rethinking the knowledge memory for lifelong model editing of large language models (Wang Peng, 2024) View paper
- Evaluation and Analysis of Sequential Editing
  - Comprehensive Evaluation Frameworks (3 papers)
  - [10] Navigating the dual facets: A comprehensive evaluation of sequential memory editing in large language models (Lin Zi-hao, 2024) View paper
  - [11] A comprehensive study of knowledge editing for large language models (Zhang, 2024) View paper
  - [31] Easyedit: An easy-to-use knowledge editing framework for large language models (Wang Peng, 2024) View paper
  - Performance Degradation Analysis (3 papers)

- [6] The butterfly effect of model editing: Few edits can trigger large language models collapse (Yang Wan-li, 2024) View paper
- [23] Understanding the limits of lifelong knowledge editing in llms (L Thede, 2025) View paper
- [47] Can we continually edit language models? on the knowledge attenuation in sequential model editing (Qi Li, 2024) View paper
- Side Effects and General Capability Preservation (1 papers)
- [13] Model editing harms general abilities of large language models: Regularization to the rescue (Jia-Chen Gu, 2024) View paper

- Contextual and Retrieval-Based Knowledge Update
  - In-Context Editing and Prompting (2 papers)
  - [1] Robust and scalable model editing for large language models (Chen Ying-fa, 2024) View paper
  - [43] Decoding by Contrasting Knowledge: Enhancing LLMs' Confidence on Edited Facts (Liu Sheng-hua, 2024) View paper
  - Retrieval-Augmented Knowledge Integration (2 papers)
  - [15] Enhancing retrieval-augmented large language models with iterative retrieval-generation synergy (Shao Zhi-hong, 2023) View paper
  - [41] Knowledge-Aware Iterative Retrieval for Multi-Agent Systems (Song Se-Young, 2025) View paper
  - Knowledge Graph-Enhanced Editing (1 papers)
  - [34] Knowledge Graph Enhanced Large Language Model Editing (Zhang Meng-qi, 2024) View paper
  - Knowledge Distillation for Update Propagation (1 papers)
  - [40] Propagating knowledge updates to lms through distillation (Padmanabhan, 2023) View paper
- Specialized Editing Applications and Extensions
  - Social Debiasing and Fairness Editing (1 papers)
  - [18] Potential and challenges of model editing for social debiasing (Yan, 2024) View paper
  - Concept-Level Knowledge Editing (1 papers)
  - [29] Editing conceptual knowledge for large language models (Wang Xiao-han, 2024) View paper
  - Unlearning as Editing (1 papers)
  - [46] Editing as Unlearning: Are Knowledge Editing Methods Strong Baselines for Large Language Model Unlearning? (Li Zexi, 2025) View paper
- Related Knowledge Update and Reasoning Paradigms
  - Continual Learning and Knowledge Expansion (3 papers)
  - [26] Refine knowledge of large language models via adaptive contrastive learning (Li Yinghui, 2025) View paper
  - [35] Bring your own knowledge: A survey of methods for llm knowledge expansion (Wang Ming-yang, 2025) View paper
  - [37] Knowledge-empowered, collaborative, and co-evolving AI models: The post-LLM roadmap (Fei Wu, 2025) View paper
  - Parameter-Efficient Fine-Tuning (1 papers)
  - [4] Step-by-step unmasking for parameter-efficient fine-tuning of large language models (Aradhye Agarwal, 2025) View paper
  - Multi-Step Reasoning and Chain-of-Thought Enhancement (7 papers)
  - [8] A survey on feedback-based multi-step reasoning for large language models on mathematics (Liu Haowei, 2025) View paper
  - [14] Improving multi-step reasoning abilities of large language models with direct advantage policy optimization (Liu Jia-cai, 2024) View paper
  - [21] Resprompt: Residual connection prompting advances multi-step reasoning in large language models (Jiang Song, 2024) View paper
  - [30] Knowledge-driven cot: Exploring faithful reasoning in llms for knowledge-intensive question answering (Wang Keheng, 2023) View paper
  - [33] Enhancing decision-making for llm agents via step-level q-value models (Zhai, 2025) View paper
  - [36] Enhancing multi-step reasoning abilities of language models through direct q-function optimization (Ji, 2024) View paper
  - [48] ART: Automatic multi-step reasoning and tool-use for large language models (Paranjape, 2023) View paper
  - Iterative Refinement and Verification (5 papers)
  - [19] Check your facts and try again: Improving large language models with external knowledge and automated feedback (Peng, 2023) View paper
  - [27] Improve: Iterative model pipeline refinement and optimization leveraging llm agents (Xue, 2025) View paper
  - [39] Pive: Prompting with iterative verification improving graph-based generative capability of llms (Han, 2024) View paper
  - [45] Iterative forward tuning boosts in-context learning in language models (Yang, 2024) View paper
  - [49] Dehallucinating large language models using formal methods guided iterative prompting (Susmit Jha, 2023) View paper
  - Knowledge Distillation and Transfer (1 papers)
  - [22] SIKeD: Self-guided Iterative Knowledge Distillation for mathematical reasoning (Adarsh Shivam, 2024) View paper
  - Sequential Instruction Tuning (2 papers)
  - [42] Fine-tuning large language models with sequential instructions (Hu, 2025) View paper
  - [50] Recursive instruction tuning of large language models for low-resource languages (Tilda Harrington, 2024) View paper
- Cross-Domain and Multimodal Applications
  - Multimodal and Speech Editing (1 papers)
  - [9] Instructspeech: Following speech editing instructions via large language models (R Huang, 2024) View paper
  - Software Engineering and Domain Modeling (2 papers)
  - [5] Multi-step iterative automated domain modeling with large language models (Yujing Yang, 2024) View paper
  - [38] From prompt design to iterative generation: Leveraging LLMs in PSE applications (Xinyu Tao, 2025) View paper
  - Scientific Discovery and Automated Benchmarking (1 papers)
  - [32] Auto-Bench: An Automated Benchmark for Scientific Discovery in LLMs (Chen Ting-ting, 2025) View paper
  - Educational Applications and Learnersourcing (2 papers)
  - [25] CIKT: A Collaborative and Iterative Knowledge Tracing Framework with Large Language Models (Li, 2025) View paper
  - [44] Exploring iterative enhancement for improving learnersourced multiple-choice question explanations with large language models (Qiming Bao, 2025) View paper
- Foundational Techniques and Surveys
  - Comprehensive Surveys and Taxonomies (2 papers)
  - [7] Knowledge Editing in Large Language Model (Javadi, 2024) View paper
  - [24] The ultimate guide to fine-tuning llms from basics to breakthroughs: An exhaustive review of technologies, research, best practices, applied research challenges and â¦ (VB Parthasarathy, 2024) View paper
  - Optimization and Training Techniques (1 papers)

- [20] Structured convergence through latent epoch reshaping for reordering intermediate computations in large language model training (Allan, 2025) View paper

Narrative

Core task: sequential model editing for large language models. The field addresses how to update factual knowledge or behavior in pretrained models through successive edits without catastrophic forgetting or performance collapse. The taxonomy reveals several main branches: Sequential Editing Methods and Architectures explores parameter-modifying techniques (including orthogonal subspace and projection-based approaches) alongside memory-augmented and meta-learning strategies; Evaluation and Analysis examines benchmarks, metrics, and the side effects of repeated edits; Contextual and Retrieval-Based Knowledge Update investigates non-parametric alternatives that store knowledge externally; Specialized Editing Applications targets domains such as debiasing or conceptual knowledge; Related Knowledge Update and Reasoning Paradigms connects editing to broader update mechanisms like iterative refinement; Cross-Domain and Multimodal Applications extends editing beyond text; and Foundational Techniques and Surveys provide overarching reviews and core methods. Representative works such as Robust Scalable Editing[1] and Wise Lifelong Editing[3] illustrate how parameter-modifying methods balance edit success with model stability, while Knowledge Editing Survey[7] offers a comprehensive landscape view.

A particularly active line of work focuses on orthogonal subspace and projection-based editing, which aims to isolate updates in low-dimensional parameter spaces to minimize interference across sequential edits. Energy Regularized Editing[0] sits squarely in this branch, proposing an energy-based regularization to preserve orthogonality and prevent gradient conflicts during lifelong editing. This contrasts with nearby approaches like O-edit Orthogonal[17], which also leverages orthogonal projections but may differ in how constraints are enforced or how edit sequences are managed. Another theme across the taxonomy is the trade-off between edit precision and generalization: some methods prioritize robust scalability (e.g., Robust Scalable Editing[1]), while others emphasize explainability or efficiency (Explainable Efficient Editing[2]). Open questions remain around how many edits a model can sustain before degradation, how to evaluate long-term side effects, and whether hybrid retrieval-augmented strategies can complement parameter updates. Energy Regularized Editing[0] contributes to the parameter-modifying paradigm by addressing energy dynamics in sequential scenarios, positioning itself among works that seek principled ways to maintain model coherence over extended edit horizons.

Related Works in Same Category

The following **1 sibling papers** share the same taxonomy leaf node with the original paper:

1. O-edit: Orthogonal subspace editing for language model sequential editing

Authors: Cai Yuchen, Cao Ding, Yuchen Cai, Ding Cao | **Year/Venue**: 2024 | **URL**: View paper

Abstract

Large language models (LLMs) acquire knowledge during pre-training, but over time, this knowledge may become incorrect or outdated, necessitating updates after training. Knowledge editing techniques address this issue without the need for costly re-training. However, most existing methods are designed for single edits, and as the number of edits increases, they often cause a decline in the model's overall performance, posing significant challenges for sequential editing. To overcome this, we pro...

Relationship Analysis

Both papers belong to the orthogonal subspace and projection-based editing category, employing projection techniques to minimize interference across sequential edits. They share the core approach of projecting edit perturbations onto orthogonal or complementary subspaces to preserve prior knowledge, with both methods analyzing weight geometry and gradient spaces. The key difference is that the original paper (SPHERE) focuses on hyperspherical energy regularization to maintain neuron uniformity on hyperspheres and projects edits away from principal hyperspherical directions, while the candidate paper (O-Edit) emphasizes gradient-based orthogonalization by projecting updates orthogonal to both previously edited knowledge gradients and implicit model knowledge gradients without the hyperspherical uniformity framework.

Contributions Analysis

Overall novelty summary. The paper introduces Hyperspherical Energy (HE) as a metric for monitoring neuron uniformity during sequential model editing and proposes SPHERE, a sparse projection method with energy regularization. It resides in the 'Orthogonal Subspace and Projection-Based Editing' leaf, which contains only two papers total. This leaf sits within the broader 'Parameter-Modifying Sequential Editing' branch, indicating a moderately sparse research direction focused on projection-based interference mitigation. The taxonomy shows that parameter-modifying methods are one of several competing paradigms, alongside parameter-preserving and retrieval-based approaches, suggesting the field is still exploring diverse architectural strategies.

The paper's leaf is adjacent to 'Neuron-Level and Layer-Targeted Editing' and 'Model Merging for Knowledge Integration' within the same parameter-modifying branch, and to 'Adapter-Based Knowledge Injection' and 'Dual-Memory Architectures' in the parameter-preserving branch. The taxonomy's scope note clarifies that orthogonal projection methods aim to prevent interference by isolating edits in complementary subspaces, distinguishing them from neuron-level targeting or external module integration. Neighboring evaluation branches examine performance degradation and side effects, indicating that stability concerns are central to the field. The paper's focus on energy dynamics connects to these evaluation themes while proposing a novel geometric lens.

Among 30 candidates examined, none clearly refute any of the three contributions. For the HE metric contribution, 10 candidates were reviewed with no refutable overlap; similarly, the theoretical proof linking HE to degradation and the SPHERE method each examined 10 candidates with zero refutations. This suggests that within the limited search scope, the specific use of hyperspherical energy for sequential editing stability appears novel. However, the small candidate pool and the presence of only one sibling paper in the taxonomy leaf mean the analysis cannot rule out related work in adjacent projection-based or energy-based editing approaches that may not have surfaced in the top-30 semantic matches.

Given the limited search scope and the sparse population of the taxonomy leaf, the paper's contributions appear relatively novel within the examined literature. The absence of refutable candidates across all three contributions, combined with the small number of sibling papers, suggests the work explores a less-crowded direction. However, the analysis is constrained by the top-30 semantic search and does not cover the full breadth of orthogonal projection or energy-based methods that may exist in the broader editing literature or related optimization domains.

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

Contribution 1: Hyperspherical Energy as a metric for sequential editing stability

Description: The authors introduce Hyperspherical Energy as a quantitative measure to assess weight uniformity throughout sequential model editing. They empirically demonstrate a strong correlation between HE dynamics and editing performance, showing that editing failures consistently coincide with uncontrolled HE fluctuations.

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. Multi-step ahead suspended sediment load modeling using machine learningâ□□multi-model approach

URL: View paper

Brief Assessment

Suspended Sediment Modeling[62] focuses on machine learning ensemble methods for sediment load prediction in hydrology. It does not address neural network weight geometry, model editing, or hyperspherical energy metrics.

---

2. Revisiting the Trade-Off Between Accuracy and Robustness via Weight Distribution of Filters

URL: View paper

Brief Assessment

Weight Distribution Robustness[68] focuses on adversarial robustness in DNNs and analyzes weight distributions between standard-trained and robust-trained models, not sequential model editing or hyperspherical energy metrics for editing stability.

---

3. Visually Explaining the Weight Distribution of Neural Networks over Time

URL: View paper

Brief Assessment

Weight Distribution Visualization[63] focuses on visualizing weight distributions over time during training, not on measuring hyperspherical uniformity for sequential model editing stability. The candidate addresses a different problem domain (visualization) rather than editing metrics.

---

4. Dynamic personalized federated learning with adaptive differential privacy

URL: View paper

Brief Assessment

Dynamic Personalized Federated[61] focuses on federated learning with differential privacy and uses Fisher information for parameter selection, not sequential model editing or hyperspherical energy metrics for editing stability.

---

5. A Novel Structure-Agnostic Multi-Objective Approach for Weight-Sharing Compression in Deep Neural Networks

URL: View paper

Brief Assessment

Structure-Agnostic Compression[65] focuses on neural network weight compression through clustering and quantization techniques for memory reduction, not on measuring weight uniformity for sequential model editing stability or knowledge retention.

---

6. The Early Phase of Neural Network Training

URL: View paper

Brief Assessment

Early Phase Training[69] focuses on the initial training phase of neural networks and weight distribution changes during early iterations, not on sequential model editing or hyperspherical energy as a stability metric for editing operations.

---

7. TRADI: Tracking deep neural network weight distributions

URL: View paper

Brief Assessment

TRADI Tracking[64] focuses on tracking weight distributions during DNN training for uncertainty estimation, not on sequential model editing or hyperspherical energy metrics for editing stability.

---

8. Adversarial Parameter Defense by Multi-Step Risk Minimization

URL: View paper

Brief Assessment

Multi-Step Risk Minimization[67] focuses on parameter corruption robustness in adversarial training contexts, not on sequential model editing or hyperspherical uniformity metrics for knowledge updates.

---

9. TRADI: Tracking deep neural network weight distributions for uncertainty estimation

URL: View paper

Brief Assessment

TRADI Uncertainty[66] focuses on tracking weight distributions during training for uncertainty estimation in DNNs, not on sequential model editing or hyperspherical energy as a stability metric for editing operations.

---

10. A theory of learning with constrained weight-distribution

URL: View paper

Brief Assessment

Constrained Weight Distribution[70] focuses on learning with prescribed weight distributions in perceptrons, not on sequential model editing or measuring editing stability. The candidate uses hyperspherical energy to analyze learning capacity under distribution constraints, whereas the original applies it to track stability during sequential editing of LLMs.

Contribution 2: Theoretical proof linking HE dynamics to knowledge degradation

Description: The authors provide a formal theoretical analysis establishing that variations in Hyperspherical Energy impose a lower bound on the interference with original pretrained knowledge. This result mathematically explains why maintaining HE stability is essential for preserving model knowledge during sequential editing.

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. Mathematical formalism for memory compression in selective state space models

URL: View paper

Brief Assessment

Memory Compression SSM[73] focuses on memory compression in selective state space models using rate-distortion theory and information-theoretic bounds. It does not address hyperspherical energy dynamics or knowledge degradation during sequential model editing, which are the core focus of the original paper's theoretical contribution.

---

2. Training in reverse: How iteration order influences convergence and stability in deep learning

URL: View paper

Brief Assessment

Training in Reverse[76] focuses on gradient update ordering and convergence stability in neural network training, not on knowledge degradation during sequential model editing or hyperspherical energy dynamics.

---

3. Incremental online learning of randomized neural network with forward regularization

URL: View paper

Brief Assessment

Incremental Online Learning[79] focuses on online learning of randomized neural networks with forward regularization for continuous model updates, not on hyperspherical energy dynamics or knowledge degradation bounds in sequential model editing contexts.

---

4. Overcoming catastrophic forgetting in neural networks

URL: View paper

Brief Assessment

Catastrophic Forgetting[77] addresses catastrophic forgetting in continual learning settings but does not provide theoretical bounds on knowledge degradation through hyperspherical energy dynamics or similar geometric measures during sequential model updates.

---

5. Forward and backward information retention for accurate binary neural networks

URL: View paper

Brief Assessment

Forward Backward Retention[72] focuses on information loss in binary neural network training through forward/backward propagation, not on hyperspherical energy dynamics or sequential model editing knowledge degradation.

---

6. The Unseen Bias: How Norm Discrepancy in Pre-Norm MLLMs Leads to Visual Information Loss

URL: View paper

Brief Assessment

Norm Discrepancy Bias[75] focuses on norm disparities in multimodal models causing visual information loss, not on hyperspherical energy dynamics during sequential model editing or knowledge degradation bounds in LLMs.

---

7. Im-loss: information maximization loss for spiking neural networks

URL: View paper

Brief Assessment

IM-loss[71] focuses on information maximization in spiking neural networks through entropy-based loss functions, not on hyperspherical energy dynamics or knowledge degradation bounds in sequential model editing. The theoretical frameworks address entirely different problems in distinct neural network paradigms.

---

8. Physics-informed neural networks for PDE problems: A comprehensive review

URL: View paper

Brief Assessment

Physics-informed Networks Review[74] focuses on physics-informed neural networks for solving partial differential equations, not on sequential model editing or knowledge degradation in language models. The candidate paper addresses completely different technical domains and does not contain relevant prior work on hyperspherical energy dynamics in model editing contexts.

---

9. Detachedly learn a classifier for class-incremental learning

URL: View paper

Brief Assessment

Detachedly Learn Classifier[80] focuses on class-incremental learning with frozen pretrained features and addresses knowledge degradation through probabilistic analysis of task-specific classifier training. It does not examine hyperspherical energy dynamics or provide theoretical bounds on knowledge degradation during sequential neural network updates as claimed in the original paper.

---

10. Lossy Loops: Shannonâ□□s DPI and Information Decay in Generative Model Training

URL: View paper

Brief Assessment

Lossy Loops[78] focuses on information-theoretic bounds in generative model training using Shannon's DPI, not on hyperspherical energy dynamics during sequential model editing.

Contribution 3: SPHERE: Sparse Projection for Hyperspherical Energy-Regularized Editing

Description: The authors introduce SPHERE, a novel regularization method that identifies a sparse space complementary to the principal hyperspherical directions of pretrained weight matrices and projects new knowledge onto it. This approach stabilizes weight distributions and preserves hyperspherical uniformity during sequential editing while maintaining general model capabilities.

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. Symphony: Edge-powered Decentralized Multi-agent Framework for Autonomous Co-evolving Intelligence

URL: View paper

Brief Assessment

Symphony Edge Framework[59] focuses on edge-powered decentralized multi-agent systems with sparse projection for parameter selection, not sequential model editing with hyperspherical energy regularization for LLMs.

---

2. Overcoming generic knowledge loss with selective parameter update

URL: View paper

Brief Assessment

Selective Parameter Update[51] focuses on continual learning for foundation models by selecting task-relevant parameters to update while preserving generic knowledge. SPHERE addresses sequential model editing with hyperspherical energy regularization to maintain weight uniformity during knowledge updates. These are distinct problem settings with different technical approaches.

---

3. Mastering Continual Reinforcement Learning through Fine-Grained Sparse Network Allocation and Dormant Neuron Exploration

URL: View paper

Brief Assessment

Sparse Network Allocation[57] focuses on continual reinforcement learning with sparse sub-network allocation for sequential task learning, not sequential model editing of language models. The technical domains and objectives are fundamentally different.

---

4. Edit Less, Achieve More: Dynamic Sparse Neuron Masking for Lifelong Knowledge Editing in LLMs

URL: View paper

Brief Assessment

Dynamic Sparse Masking[58] focuses on neuron-level masking via entropy-guided selection to identify knowledge-general and knowledge-specific neurons for lifelong editing. SPHERE targets sparse projection onto spaces complementary to principal hyperspherical directions to preserve hyperspherical uniformity. These are distinct technical approaches with different theoretical foundations and mechanisms.

---

5. DySK-Attn: A Framework for Efficient, Real-Time Knowledge Updating in Large Language Models via Dynamic Sparse Knowledge Attention

URL: View paper

Brief Assessment

DySK-Attn[55] focuses on dynamic knowledge graph integration via sparse attention mechanisms for real-time knowledge updating, not on sequential model editing or hyperspherical energy regularization of weight matrices.

---

6. RoseLoRA: Row and Column-wise Sparse Low-rank Adaptation of Pre-trained Language Model for Knowledge Editing and Fine-tuning

URL: View paper

Brief Assessment

RoseLoRA[56] focuses on parameter-efficient fine-tuning through row and column-wise sparse low-rank adaptation for knowledge editing, not on hyperspherical energy regularization or preserving hyperspherical uniformity during sequential editing.

---

7. Symphony: A Decentralized Multi-Agent System for Co-Evolving Intelligence at Scale

URL: View paper

Brief Assessment

Symphony Decentralized[60] focuses on decentralized multi-agent systems with sparse projection for parameter selection in distributed learning contexts, not on sequential model editing or hyperspherical energy regularization for knowledge preservation in LLMs.

---

8. Continual Learning via Sparse Memory Finetuning

URL: View paper

Brief Assessment

Sparse Memory Finetuning[54] focuses on continual learning in language models by updating sparse memory slots to prevent catastrophic forgetting, not on sequential model editing with hyperspherical energy regularization as in SPHERE.

---

9. Robust Learning of Diverse Code Edits

URL: View paper

Brief Assessment

Diverse Code Edits[53] focuses on code editing tasks using sparse projection to preserve base model capabilities during fine-tuning, not on sequential knowledge editing in LLMs or hyperspherical energy regularization for weight uniformity.

---

10. Model Unlearning via Sparse Autoencoder Subspace Guided Projections

URL: View paper

Brief Assessment

Sparse Autoencoder Unlearning[52] focuses on machine unlearning (removing specific knowledge) using SAE-guided subspace projections, while SPHERE addresses sequential model editing (updating knowledge) using hyperspherical energy regularization. These are fundamentally different tasks with different objectives and methodologies.

Appendix: Text Similarity Detection

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

References

- [0] Energy-Regularized Sequential Model Editing on Hyperspheres View paper
- [1] Robust and scalable model editing for large language models View paper
- [2] Explainable and efficient editing for large language models View paper
- [3] Wise: Rethinking the knowledge memory for lifelong model editing of large language models View paper
- [4] Step-by-step unmasking for parameter-efficient fine-tuning of large language models View paper
- [5] Multi-step iterative automated domain modeling with large language models View paper
- [6] The butterfly effect of model editing: Few edits can trigger large language models collapse View paper
- [7] Knowledge Editing in Large Language Model View paper
- [8] A survey on feedback-based multi-step reasoning for large language models on mathematics View paper
- [9] Instructspeech: Following speech editing instructions via large language models View paper
- [10] Navigating the dual facets: A comprehensive evaluation of sequential memory editing in large language models View paper
- [11] A comprehensive study of knowledge editing for large language models View paper
- [12] Selective Knowledge Injection via Adapter Modules in Largeâ□□Scale Language Models View paper
- [13] Model editing harms general abilities of large language models: Regularization to the rescue View paper
- [14] Improving multi-step reasoning abilities of large language models with direct advantage policy optimization View paper
- [15] Enhancing retrieval-augmented large language models with iterative retrieval-generation synergy View paper
- [16] Model Merging for Knowledge Editing View paper

- [17] O-edit: Orthogonal subspace editing for language model sequential editing View paper
- [18] Potential and challenges of model editing for social debiasing View paper
- [19] Check your facts and try again: Improving large language models with external knowledge and automated feedback View paper
- [20] Structured convergence through latent epoch reshaping for reordering intermediate computations in large language model training View paper
- [21] Resprompt: Residual connection prompting advances multi-step reasoning in large language models View paper
- [22] SIKeD: Self-guided Iterative Knowledge Distillation for mathematical reasoning View paper
- [23] Understanding the limits of lifelong knowledge editing in llms View paper
- [24] The ultimate guide to fine-tuning llms from basics to breakthroughs: An exhaustive review of technologies, research, best practices, applied research challenges and â View paper
- [25] CIKT: A Collaborative and Iterative Knowledge Tracing Framework with Large Language Models View paper
- [26] Refine knowledge of large language models via adaptive contrastive learning View paper
- [27] Improve: Iterative model pipeline refinement and optimization leveraging llm agents View paper
- [28] Neuron-Level Sequential Editing for Large Language Models View paper
- [29] Editing conceptual knowledge for large language models View paper
- [30] Knowledge-driven cot: Exploring faithful reasoning in llms for knowledge-intensive question answering View paper
- [31] Easyedit: An easy-to-use knowledge editing framework for large language models View paper
- [32] Auto-Bench: An Automated Benchmark for Scientific Discovery in LLMs View paper
- [33] Enhancing decision-making for llm agents via step-level q-value models View paper
- [34] Knowledge Graph Enhanced Large Language Model Editing View paper
- [35] Bring your own knowledge: A survey of methods for llm knowledge expansion View paper
- [36] Enhancing multi-step reasoning abilities of language models through direct q-function optimization View paper
- [37] Knowledge-empowered, collaborative, and co-evolving AI models: The post-LLM roadmap View paper
- [38] From prompt design to iterative generation: Leveraging LLMs in PSE applications View paper
- [39] Pive: Prompting with iterative verification improving graph-based generative capability of llms View paper
- [40] Propagating knowledge updates to lms through distillation View paper
- [41] Knowledge-Aware Iterative Retrieval for Multi-Agent Systems View paper
- [42] Fine-tuning large language models with sequential instructions View paper
- [43] Decoding by Contrasting Knowledge: Enhancing LLMs' Confidence on Edited Facts View paper
- [44] Exploring iterative enhancement for improving learnersourced multiple-choice question explanations with large language models View paper
- [45] Iterative forward tuning boosts in-context learning in language models View paper
- [46] Editing as Unlearning: Are Knowledge Editing Methods Strong Baselines for Large Language Model Unlearning? View paper
- [47] Can we continually edit language models? on the knowledge attenuation in sequential model editing View paper
- [48] ART: Automatic multi-step reasoning and tool-use for large language models View paper
- [49] Dehallucinating large language models using formal methods guided iterative prompting View paper
- [50] Recursive instruction tuning of large language models for low-resource languages View paper
- [51] Overcoming generic knowledge loss with selective parameter update View paper
- [52] Model Unlearning via Sparse Autoencoder Subspace Guided Projections View paper
- [53] Robust Learning of Diverse Code Edits View paper
- [54] Continual Learning via Sparse Memory Finetuning View paper
- [55] DySK-Attn: A Framework for Efficient, Real-Time Knowledge Updating in Large Language Models via Dynamic Sparse Knowledge Attention View paper
- [56] RoseLoRA: Row and Column-wise Sparse Low-rank Adaptation of Pre-trained Language Model for Knowledge Editing and Fine-tuning View paper
- [57] Mastering Continual Reinforcement Learning through Fine-Grained Sparse Network Allocation and Dormant Neuron Exploration View paper
- [58] Edit Less, Achieve More: Dynamic Sparse Neuron Masking for Lifelong Knowledge Editing in LLMs View paper
- [59] Symphony: Edge-powered Decentralized Multi-agent Framework for Autonomous Co-evolving Intelligence View paper
- [60] Symphony: A Decentralized Multi-Agent System for Co-Evolving Intelligence at Scale View paper
- [61] Dynamic personalized federated learning with adaptive differential privacy View paper
- [62] Multi-step ahead suspended sediment load modeling using machine learningâmulti-model approach View paper
- [63] Visually Explaining the Weight Distribution of Neural Networks over Time View paper
- [64] TRADI: Tracking deep neural network weight distributions View paper
- [65] A Novel Structure-Agnostic Multi-Objective Approach for Weight-Sharing Compression in Deep Neural Networks View paper
- [66] TRADI: Tracking deep neural network weight distributions for uncertainty estimation View paper
- [67] Adversarial Parameter Defense by Multi-Step Risk Minimization View paper
- [68] Revisiting the Trade-Off Between Accuracy and Robustness via Weight Distribution of Filters View paper
- [69] The Early Phase of Neural Network Training View paper
- [70] A theory of learning with constrained weight-distribution View paper
- [71] Im-loss: information maximization loss for spiking neural networks View paper
- [72] Forward and backward information retention for accurate binary neural networks View paper
- [73] Mathematical formalism for memory compression in selective state space models View paper
- [74] Physics-informed neural networks for PDE problems: A comprehensive review View paper
- [75] The Unseen Bias: How Norm Discrepancy in Pre-Norm MLLMs Leads to Visual Information Loss View paper
- [76] Training in reverse: How iteration order influences convergence and stability in deep learning View paper
- [77] Overcoming catastrophic forgetting in neural networks View paper
- [78] Lossy Loops: Shannonâs DPI and Information Decay in Generative Model Training View paper
- [79] Incremental online learning of randomized neural network with forward regularization View paper
- [80] Detachedly learn a classifier for class-incremental learning View paper