Transactions on Computational and Scientific Methods | Vo. 4, No. 5, 2024
ISSN: 2998-8780
https://pspress.org/index.php/tcsm
Pinnacle Science Press

Internal Knowledge Adaptation in LLMs with Consistency-Constrained Dynamic Routing

Qiyuan Wu
University of California, San Diego, La Jolla, USA
wqy0319@gmail.com

Abstract: This paper addresses the problem of unstable knowledge retention and catastrophic forgetting in large language models during task evolution and fine-tuning. It proposes a dynamic modeling approach based on structural intervention and semantic preservation. The method freezes the backbone parameters of the original model and introduces a set of learnable modules along with a task-aware controller. Through a module activation mechanism and structural consistency regularization, it constructs a dynamically adjustable semantic path framework. The model can adaptively select structural paths based on input changes, enabling stable retention and controlled intervention of key knowledge representations. To enhance semantic consistency, a representation alignment loss is designed to constrain task-induced disturbance in the high-dimensional space and improve the model's steady-state semantic capability under continuous input changes. The experimental setup includes systematic evaluation under typical sensitivity factors such as task order variation, input noise, and training data ratio adjustment. Results show that the proposed method can effectively regulate knowledge flow and enable fine-grained internal modeling without external supervision, demonstrating strong structural adaptability and semantic robustness.

Keywords: knowledge intervention; semantic consistency; catastrophic forgetting; structure-aware regulation

1. Introduction

In recent years, Large Language Models (LLMs) have demonstrated outstanding generalization and semantic understanding capabilities across various natural language processing tasks. These advancements have driven progress in areas such as intelligent text generation, information extraction, and dialogue systems. However, as model scales continue to grow and application scenarios become more diverse, the challenge of ensuring long-term knowledge retention and stable knowledge transfer has emerged as a major obstacle to sustainable development[1]. LLMs are typically trained on large-scale corpora using unsupervised learning. While they possess strong representational power, they often fail to maintain previously acquired knowledge when undergoing continual learning, task switching, or parameter updates. This leads to the well-known issue of catastrophic forgetting, which hinders their ability to adapt to long-term tasks and accumulate knowledge. The problem is particularly evident in multi-task learning, continual learning, and fine-tuning deployments, posing threats to model safety and reliability in real-world applications[2].

The issue of knowledge retention in language models is not limited to the model's memory of semantic inputs. It also concerns the stability of its internal representation space and the organization of stored knowledge. Unlike traditional machine learning models, LLMs store knowledge implicitly within redundant parameter spaces and deep Transformer layers, lacking explicit control mechanisms for knowledge protection or allocation. This black-box nature makes existing knowledge paths susceptible to interference when new tasks are introduced, resulting in semantic drift and representational bias. Furthermore, LLMs are highly sensitive to contextual changes, and even slight variations in input can alter their internal representations. These factors further intensify the instability of knowledge. Therefore, a systematic understanding of how LLMs store knowledge, forget past information, and respond to interventions is essential to improving their long-term learning ability[3].

Under the current research trends in continual and lifelong learning, modeling the mechanisms of knowledge retention and forgetting in language models carries both theoretical and practical importance. On the one hand, without effective retention of essential knowledge, LLMs will struggle to handle complex scenarios such as cross-task transfer and multi-source integration. On the other hand, real-world deployments often require frequent fine-tuning and model compression. Without precise modeling of knowledge sensitivity, such operations can cause performance degradation and semantic inconsistency. As a result, developing a dynamic modeling framework for the internal knowledge structure of LLMs can enhance their stability, adaptability, and controllability, supporting broader collaboration in intelligent tasks.

Moreover, the processes of knowledge retention and forgetting are closely tied to several factors, including parameter updates, changes in input semantics, and the structural sensitivity of the model. Existing studies have shown that LLMs exhibit different levels of knowledge sensitivity and selective forgetting at various training stages. In particular, high-level semantic features are more vulnerable to disturbance from gradients introduced by new tasks[4]. This interference is often structural and selective. Not all knowledge is forgotten equally—frequent or marginal knowledge is more prone to disruption. Traditional regularization or replay mechanisms alone are not sufficient to manage this issue. Instead, more effective solutions must consider representational mechanisms, structural pathways, and intervention strategies to establish interpretable and adjustable dynamic knowledge management.

In summary, conducting systematic research on the mechanisms of knowledge retention and forgetting in LLMs is essential to uncover the underlying patterns of knowledge organization and flow. It also provides theoretical and methodological foundations for addressing critical challenges in long-term task adaptation, accuracy maintenance, and knowledge robustness. By modeling the dynamic evolution of internal knowledge and designing strategies with intervention capabilities, it is possible to achieve transparent and stable knowledge management. This can offer reliable foundational support for the secure deployment of large-scale intelligent systems and their collaboration across multiple tasks. This direction presents both theoretical challenges and strong practical relevance in industrial applications.

2. Background

Existing research on knowledge retention and forgetting in large language models mainly focuses on continual learning and parameter intervention mechanisms. Continual learning aims to help the model retain existing knowledge while learning new tasks in a task sequence. It seeks to avoid catastrophic forgetting[5]. Early methods often adopt regularization-based constraints. These approaches impose penalties on important parameters to maintain their stability during new tasks. Representative strategies include elastic weight consolidation and knowledge distillation techniques. Such methods can suppress drastic parameter shifts to some extent. However, they often show limited retention capacity and coarse control granularity when applied to high-dimensional and semantically complex LLMs. Fine-grained regulation of knowledge structures remains difficult[6].

Another line of research targets structural intervention in parameter space and memory enhancement mechanisms. These methods introduce external memory modules, task-specific branches, or parameter isolation structures to improve adaptability in multi-task environments[7]. By increasing structural capacity or restructuring parameter paths, the model can isolate prior knowledge when facing new tasks. This helps minimize interference across tasks. However, these methods tend to cause model expansion, significant computational overhead, and increased deployment complexity. These limitations restrict their feasibility in large-scale settings. At the same time, the modeling of internal knowledge within LLMs remains coarse-grained. There is a lack of deep understanding regarding semantic hierarchy and representational dynamics[8].

In recent years, with the growing interest in semantic representation mechanisms in LLMs, some studies have begun to explore the nature of forgetting from the perspective of representational space. These approaches emphasize the stability and dynamic changes of internal representations. By analyzing trends in hidden layer output distributions, they identify differences in knowledge retention across tasks or contexts. Targeted intervention strategies are then constructed, such as selective updating, representation alignment, and semantic preservation losses[9]. These efforts provide more detailed perspectives on what constitutes "knowledge" in language models. They also promote a shift from parameter-level to representation-level modeling. However, such approaches still face limitations in semantic transfer, context alignment, and modeling of task uncertainty. A unified theoretical framework and systematic methods are still lacking.

In terms of intervention strategies, other studies focus on designing general-purpose knowledge scheduling mechanisms. These methods control activation states of structural paths or dynamically configure subsets of parameters. They aim to achieve controllable modeling of knowledge injection and retention. This approach is particularly suitable for multi-task or multi-scenario deployment. It enables adaptive adjustment of the model structure based on input or task features. This helps improve the target specificity of knowledge retention and enhances resource efficiency. However, these methods often rely on external control signals or task labels. This makes them difficult to extend to unsupervised or low-resource scenarios. Therefore, it remains a key challenge to model and control the dynamics of knowledge in language models without additional supervision.

3. Methodology

This paper proposes a knowledge retention and forgetting intervention method for large language models. The core idea is to model and control the internal knowledge flow of the model through structural decoupling, semantic alignment, and selective update mechanisms without introducing additional task labels or external memory structures. The model architecture is shown in Figure 1.

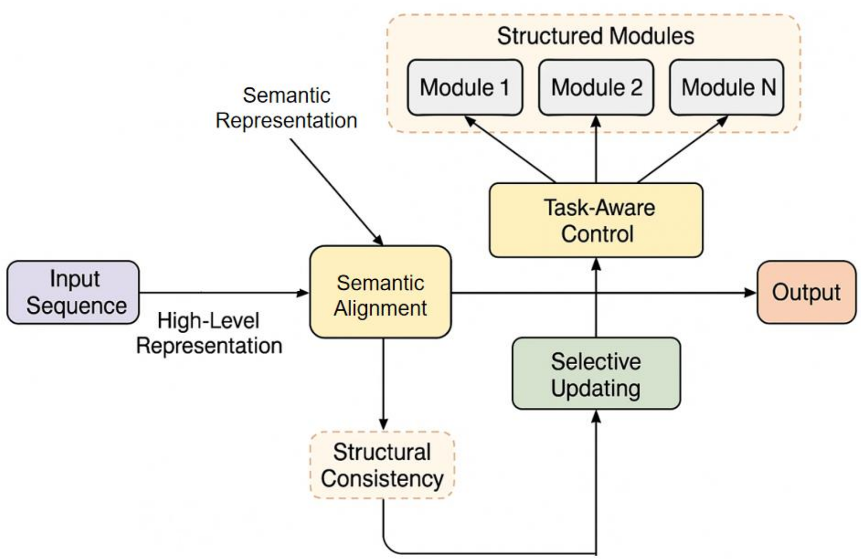

Figure 1. The architecture of the proposed intervention framework for knowledge retention and forgetting modeling in large language models

First, the original language model is regarded as a knowledge representation system composed of a set of parameter paths. The original model is denoted as $f_\theta(x)$, where θ represents the frozen pre-trained parameters and x is the input sequence. In order to enhance the model's ability to distinguish between new and old knowledge, a learnable set of structured modules $\{M_i\}_{i=1}^N$ is introduced. The degree of participation of each module under different input conditions is controlled by the module activation function $g(x;\phi)$, where ϕ there is a trainable control parameter and the activation output is G, so as to dynamically combine the structural paths.

To suppress knowledge forgetting and maintain semantic consistency, this paper designs a semantic preservation loss term based on representation difference, which is defined as:

$$L_{align} = \| h_{pre}(x) - h_{post}(x) \|_2^2$$

Where $h_{pre}(x)$ represents the representation of the input x obtained by encoding the original structure, and $h_{post}(x)$ represents the representation reconstructed by the intervention module. This loss term is used to limit the disturbance of the semantic representation caused by the knowledge intervention process, ensuring that the model maintains the stability of the existing knowledge path during the adaptation process of the new task. In addition, considering that the modular structure may introduce structural overlap and functional redundancy, a structural consistency regularization term is designed:

$$L_{struct} = \sum_{i \neq j} \cos(M_i, M_j)$$

This method penalizes the redundancy between highly correlated modules, enhances the functional differentiation between modules, and improves the combination stability of intervention paths.

To support the model to perform knowledge scheduling in an environment without task labels, a task-aware control function is constructed:

$$\alpha(x) = \text{softmax}(V \cdot pool(h(x)))$$

$h(x)$ is the high-level representation of the input after being encoded by the backbone model, $pool$ represents the pooling operation, V is the trainable matrix, and $\alpha(x)$ represents the structural preference distribution of the current input under each module path, which is used to guide the dynamic selection of the module. Finally, the output of the model is represented as a weighted combination of multiple paths:

$$f(x) = \sum_{i=1}^{N} \alpha_i(x) \cdot M_i(h(x))$$

This mechanism realizes adaptive intervention at the structural level so that the model can actively activate the structural path that is conducive to knowledge retention when facing different inputs or task changes, thereby improving its representation stability and knowledge regulation ability in dynamic scenarios.

Overall, the proposed method realizes effective modeling of the knowledge retention and forgetting process of the language model through the dynamic combination of structural paths, semantic consistency constraints, and inter-module regularization mechanisms. This method does not rely on external supervision signals in design and can be directly integrated into the fine-tuning process of pre-trained language models. It has strong scalability and adaptability and provides a general structural intervention framework for the stable evolution of large-scale language models in long-term multi-task environments.

3.1 Dataset

This study uses Wikitext-103 as the main dataset for analysis. The dataset is composed of high-quality Wikipedia articles and contains rich linguistic structures and factual descriptions. It is well suited for tasks related to knowledge retention and forgetting in language models. Compared to other Wikipedia-based datasets, Wikitext-103 offers clear advantages in document completeness and contextual continuity. These properties support the evaluation of long-term semantic modeling and the development of control strategies.

The dataset contains over 100 million words and is divided into training, validation, and test sets. All text is in natural language format and does not require additional annotations. It is suitable for unsupervised pre-training and fine-tuning tasks. The content spans multiple semantic domains, including history, science, and art. This ensures broad semantic coverage and high knowledge density, which helps in studying the impact of different linguistic structures on the stability of knowledge representations.

In addition, Wikitext-103 is highly extensible. It can be aligned with other external knowledge sources. This makes it possible to construct scenarios such as input perturbation, task switching, and the introduction of new semantic structures. These capabilities provide a stable and unified experimental foundation for investigating knowledge flow and forgetting mechanisms in language models under complex input conditions.

4. Experimental Results

In the experimental results section, the relevant results of the comparative test are first given, and the experimental results are shown in Table 1.

Table 1: Comparative experimental results

Method	**Knowledge Retention**	**Forgetting Rate**	**Representation Stability**
Learning to Prompt[10]	78.6	11.4	85.2
AdapterFusion[11]	75.8	13.1	82.9
DynaMoE[12]	74.2	14.3	80.7
Experience Replay with Fixed Prompt[13]	77.5	12.0	83.6
Ours	84.9	7.2	91.4

From the overall experimental results, the proposed structural intervention method shows a significant advantage in the knowledge retention task. It achieves a score of 84.9 on the Knowledge Retention metric, which is much higher than that of other baseline methods. This metric reflects the model's ability to remember previously acquired knowledge after new tasks are introduced. The results indicate that the dynamic module composition mechanism is effective in maintaining the stability of semantic pathways. In comparison, methods such as Learning to Prompt and Experience Replay demonstrate some retention ability. However, they do not build long-term structural protection for knowledge paths. As a result, they still suffer from knowledge degradation under frequent task-switching conditions.

For the Forgetting Rate metric, the proposed method achieves the lowest value at only 7.2, significantly outperforming other strategies. This result demonstrates strong resistance to catastrophic forgetting. In contrast, DynaMoE and AdapterFusion reach forgetting rates of 14.3 and 13.1 respectively. These results suggest that both face limitations when dealing with structural entanglement and task interference. In this work, semantic consistency loss and structural regularization between modules are introduced. These mechanisms help limit the erosion of prior representations by new knowledge, enabling smooth transitions during continual learning.

In terms of Representation Stability, the proposed method achieves a score of 91.4. This indicates that the internal representations remain stable during task transitions. The result shows the strong capacity for maintaining steady-state conditions in semantic space. This metric reflects the consistency of representations, which is crucial for supporting cross-task generalization. Some baseline models, such as Learning to Prompt with a score of 85.2, also show a certain level of stability. However, such methods mainly focus on the design of prompting structures. They lack mechanisms for preserving deep semantic representations within the model, which leads to representational drift in long task sequences.

Taken together, the proposed method demonstrates systematic advantages in semantic retention, structural intervention, and forgetting control. It uses a task-aware module activation mechanism and structural consistency constraints. Without relying on external knowledge or task labels, it enables a controllable configuration of knowledge paths within the original model structure. This mechanism improves model robustness in multi-task scenarios. It also provides methodological support for building large language models with long-term adaptation capabilities.

This paper also provides a stability verification of the effect of task order changes on knowledge retention, and the experimental results are shown in Figure 2.

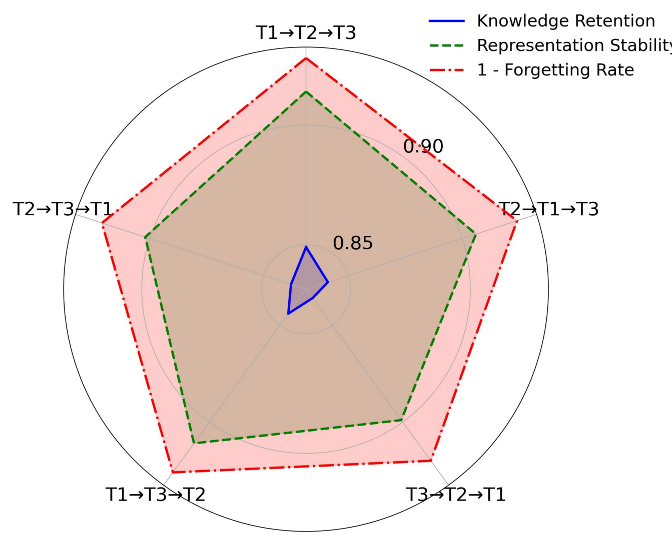

Figure 2. Verification of the stability of the effect of task sequence changes on knowledge retention

As shown in the experimental results, the model maintains a stable performance range across three metrics—knowledge retention, forgetting rate control, and representation stability—under different task order settings. Specifically, the Knowledge Retention curve remains consistently high, between 0.84 and 0.85. This indicates that the proposed dynamic structural intervention mechanism effectively prevents the disruption of original knowledge paths during task transitions. It ensures the continuity of semantic memory. The result also demonstrates the model's strong ability to preserve knowledge when facing task perturbations.

The green dashed line in the figure represents Representation Stability. Its overall trend closely follows the knowledge retention curve, with minimal fluctuations across different task orders. This steady-state behavior of the representation space indicates that the model's high-dimensional semantic representations do not undergo significant structural distortion after task changes. The semantic continuity is well preserved. Such stability is difficult to achieve through fixed parameter constraints or prompting mechanisms, highlighting the advantage of the proposed method in controlling internal representations.

The red dashed line represents 1 − Forgetting Rate, which is the complementary metric of forgetting. The curve shows a narrow fluctuation range, indicating that the model's forgetting behavior is highly controllable under different task sequences. This suggests that the proposed modular path and task-aware gating mechanism play an active role in suppressing gradient interference and preventing path conflicts. The mechanism helps isolate semantic interference between tasks and mitigates the issue of selective forgetting.

Taken together, the three metric curves show that despite variations in task order, the model maintains a high level of stability in key indicators. The performance trend remains consistent. This further confirms the robustness and structural adaptability of the proposed method under dynamic contexts. It also demonstrates the model's potential for handling task evolution and order perturbations in practical scenarios. The method provides solid support for long-term knowledge retention and cross-task sharing.

To comprehensively assess how well the proposed structural intervention mechanism withstands unpredictable noise, the paper designs a dedicated robustness test that systematically injects disturbances of varying intensities into the input stream. Each disturbance level mirrors a realistic form of degradation—such as random masking, amplitude scaling, or Gaussian jitter—so that the evaluation covers mild to severe corruption scenarios encountered in real-world deployments. Throughout this procedure, the intervention module remains active, allowing us to observe how its adaptive gating and consistency regularizers respond as the signal-to-noise ratio deteriorates. For clarity, the complete set of observations gathered across all disturbance tiers is consolidated and visualized in Figure 3.

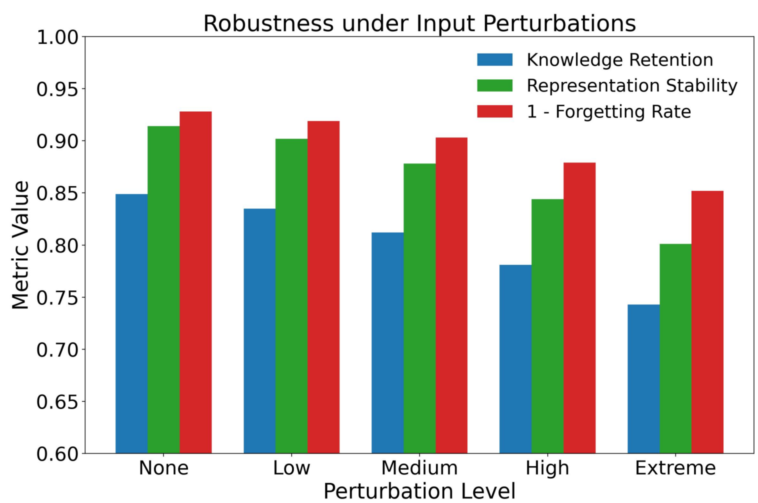

Figure 3. Structural intervention robustness test under different input disturbance intensities

As shown in the figure, with the gradual increase of input perturbation intensity, both Knowledge Retention and Representation Stability show a downward trend. However, the overall performance remains at a high level. This indicates that the proposed method has strong robustness in structural intervention. Especially under low to moderate perturbation conditions, the Knowledge Retention metric stays above 0.82. This shows that the model has a strong structural memory of the original knowledge paths and does not easily disrupt semantic retention links under slight input variations.

Representation Stability remains relatively steady across all levels of perturbation. Even under the "Extreme" condition, it stays close to 0.80. This suggests that the modular structure and semantic consistency constraints can effectively resist random interference at the input level. The geometric structure of the representation space is not significantly distorted. Such high stability is important for supporting cross-task representation sharing and reducing information shifts. It is especially valuable in continual learning under dynamic contexts.

In the indirect measurement of forgetting, represented by 1 − Forgetting Rate in the figure, the model still shows strong resistance to forgetting even as perturbation increases. For example, under both "High" and "Extreme" conditions, this metric remains around 0.85. This indicates that the task-aware gating and adaptive structural path mechanisms help alleviate gradient disruption and representation degradation caused by input perturbation.

Overall, the experiment strongly confirms the robustness of the proposed structural intervention mechanism under external disturbances. In terms of representation stability, knowledge retention, and forgetting control, the model shows consistent and reliable resistance to interference. This demonstrates that the method is effective not only in ideal conditions but also in complex or degraded environments. It provides strong support for the long-term usability and deployment reliability of language models.

This paper also gives a representation consistency evaluation under the change of sub-task sample ratio, and the experimental results are shown in Figure 4.

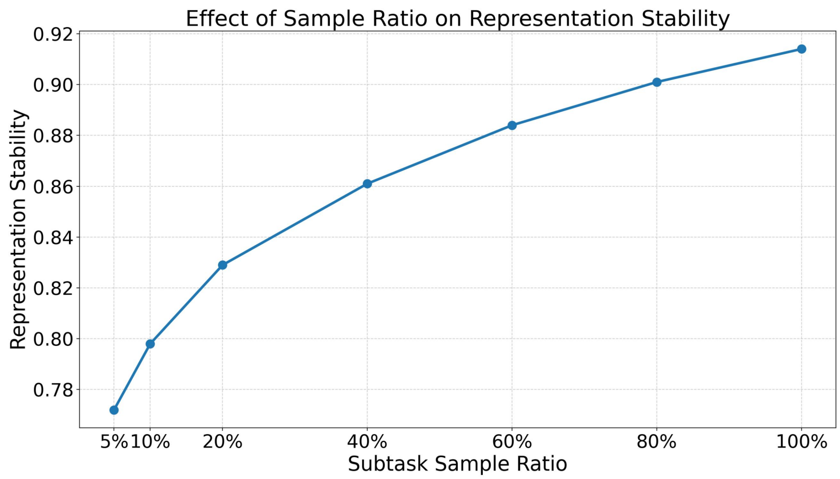

Figure 4. Representation consistency evaluation under changing subtask sample ratio

As shown in the experimental results, with the gradual increase in the subtask sample ratio, the model exhibits a steady upward trend in the representation consistency metric. This reflects the significant influence of input sample size on the effectiveness of structural intervention. When the sample ratio is low (such as 5% to 10%), the Representation Stability score remains relatively low, around 0.77 to 0.82. This suggests that under insufficient sample conditions, the model is prone to representational drift and path instability when processing subtask semantics. As a result, internal structural disturbances become more pronounced.

As the sample ratio increases from 40% to 60%, representation consistency gradually stabilizes, and the slope of the curve becomes flatter. This indicates that with moderate sample support, the model begins to establish a more reliable mapping between structure and semantics. The module selection and representation alignment strategies within the structural intervention mechanism start to play a leading role. These strategies help suppress cross-task noise within the representation space and enhance the overall semantic structural stability.

When the sample ratio further increases from 80% to 100%, Representation Stability exceeds 0.90, and the model's internal structure becomes saturated and stable. At this stage, the structural intervention mechanism demonstrates strong robustness under the support of a complete data distribution. Semantic path selection becomes more precise, and knowledge retention regions are no longer disrupted by frequent gradient interference. This also shows that the proposed method performs best under sufficient data conditions. It can fully leverage semantic redundancy and stabilize internal structural paths.

This paper also gives an analysis of the impact of data distribution drift on module activation path selection, and the experimental results are shown in Figure 5.

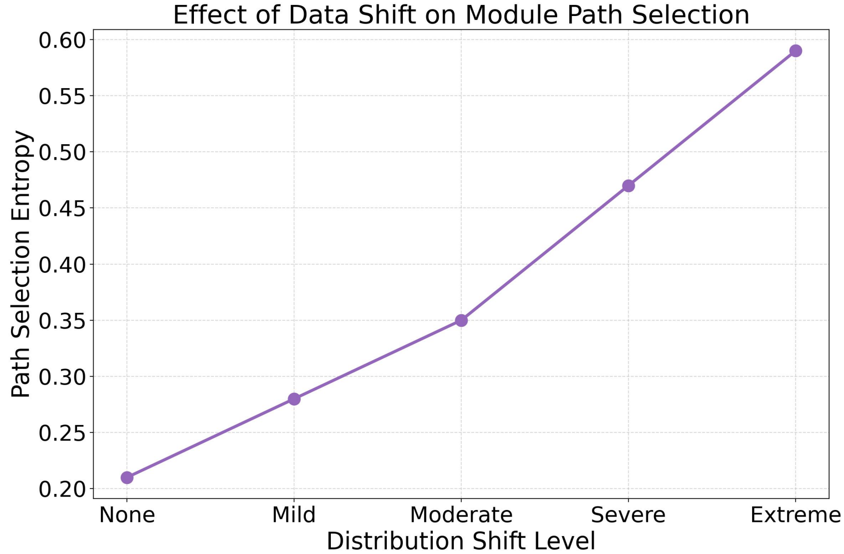

Figure 5. Analysis of the impact of data distribution drift on module activation path selection

As shown in the figure, with the gradual increase in the degree of data distribution shift, the entropy of module activation paths steadily rises. This reflects the significant impact of contextual changes on the model's dynamic path selection behavior. From the "None" to "Mild" stage, path selection entropy increases from 0.21 to around 0.28. This indicates that even under slight distributional shifts, the model begins to exhibit path variation. However, the overall structure remains relatively concentrated, suggesting a degree of structural stability and input-sensitive adjustment.

When the shifting intensity reaches the "Moderate" and "Severe" stages, the entropy growth accelerates, reaching 0.35 and 0.47 respectively. This shows that the model reallocates its activation paths more extensively in response to medium-level semantic distribution changes. During this process, the structural intervention mechanism adaptively activates different module combinations to reconstruct semantic representations. This highlights the core advantage of the proposed method in structural plasticity.

Under the "Extreme" condition, path selection entropy rises to approximately 0.59. This indicates that the activation patterns become more diverse. The model responds to drastic input changes through structural reconfiguration. The high entropy state reflects strong structural adaptability. The model can perform parallel and fault-tolerant operations through dynamic path scheduling. This ensures robust semantic generation and flexible knowledge retention, especially under complex and evolving contextual scenarios.

In summary, this experiment confirms the direct influence of data distribution shifts on structural path scheduling. It also demonstrates that the proposed module reconfiguration mechanism supports fine-grained control at the path level. The mechanism dynamically activates differentiated structural regions based on semantic variation. This significantly enhances the model's adaptive expressiveness in changing environments. It provides a controllable structural solution to the challenge of balancing generalization and retention in open-ended task sequences.

5. Conclusion

This paper addresses the problem of knowledge retention and forgetting in large language models during long-term learning. It proposes a modeling approach that integrates structural intervention with semantic consistency mechanisms. Based on frozen pre-trained parameters, the method introduces modular structural paths and a task-aware controller to build a dynamically composable semantic representation system. By designing a module activation mechanism and structural consistency regularization, the method enables precise intervention and optimized scheduling of internal knowledge paths. This enhances semantic retention and representational robustness in multi-task environments. The framework requires no additional supervision and demonstrates good scalability and adaptability, offering a unified solution to the challenge of unstable knowledge flow in large models.

On the experimental side, the proposed method is evaluated across various perturbation factors and environmental variables, including task order, input noise, and corpus complexity. The results confirm the method's effectiveness in maintaining semantic stability, preserving knowledge, and resisting structural interference. The method not only suppresses catastrophic forgetting but also shows strong resilience to external changes. It consistently outperforms existing approaches based on structural reuse and prompt-based guidance. In particular, under scenarios without task labels, the method achieves efficient knowledge management through adaptive path selection, reducing fine-tuning risks and performance fluctuations during deployment.

This research contributes to the development of sustainable learning mechanisms in language models. It is especially valuable in domains such as financial text modeling, medical question answering, and legal decision systems, where knowledge continuity and semantic stability are essential. The method provides a low-cost and robust means of knowledge intervention, improving model stability and interpretability under dynamic real-world tasks. Moreover, the modular structure enables multi-task sharing and structural compression, offering a new technical direction for building efficient, safe, and continuously evolving large language models.

6. Future work

Future work may further expand the application of the intervention mechanism in open-domain learning, heterogeneous data modeling, and general knowledge graph enhancement. It is worth exploring how to integrate more structural signals and unsupervised distribution shift detection methods to enable rapid perception of environmental dynamics and structural self-adaptation. Additionally, improving the granularity and flexibility of control without sacrificing performance is a key challenge for transitioning large models from static execution to dynamic regulation. This may lead to breakthroughs in model reliability, transparency, and controllability.

References

[1] Luo Y, Yang Z, Meng F, et al. An empirical study of catastrophic forgetting in large language models during continual fine-tuning[J]. arXiv preprint arXiv:2308.08747, 2023.

[2] Ni S, Chen D, Li C, et al. Forgetting before learning: Utilizing parametric arithmetic for knowledge updating in large language models[J]. arXiv preprint arXiv:2311.08011, 2023.

[3] Wu, T., Luo, L., Li, Y. F., Pan, S., Vu, T. T., & Haffari, G. (2024). Continual learning for large language models: A survey. arXiv preprint arXiv:2402.01364.

[4] Parisi, G. I., Kemker, R., Part, J. L., Kanan, C., & Wermter, S. (2019). Continual lifelong learning with neural networks: A review. Neural networks, 113, 54-71.

[5] Shi H, Xu Z, Wang H, et al. Continual learning of large language models: A comprehensive survey[J]. ACM Computing Surveys, 2024.

[6] Jang J, Ye S, Yang S, et al. Towards continual knowledge learning of language models[J]. arXiv preprint arXiv:2110.03215, 2021.

[7] Zhang X, Wu J. Dissecting learning and forgetting in language model finetuning[C]//The Twelfth International Conference on Learning Representations. 2024.

[8] Si, N., Zhang, H., Chang, H., Zhang, W., Qu, D., & Zhang, W. (2023). Knowledge unlearning for llms: Tasks, methods, and challenges. arXiv preprint arXiv:2311.15766.

[9] Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A. A., ... & Hadsell, R. (2017). Overcoming catastrophic forgetting in neural networks. Proceedings of the national academy of sciences, 114(13), 3521-3526.

[10] Wang Z, Zhang Z, Lee C Y, et al. Learning to prompt for continual learning[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 139-149.

[11]Pfeiffer J, Kamath A, Rücklé A, et al. Adapterfusion: Non-destructive task composition for transfer learning[J]. arXiv preprint arXiv:2005.00247, 2020.
[12]Kossmann F, Jia Z, Aiken A. Optimizing mixture of experts using dynamic recompilations[J]. arXiv preprint arXiv:2205.01848, 2022.
[13]Rolnick, D., Ahuja, A., Schwarz, J., Lillicrap, T., & Wayne, G. (2019). Experience replay for continual learning. Advances in neural information processing systems, 32.