INTERNATIONAL JOURNAL OF ADVANCED ARTIFICIAL INTELLIGENCE RESEARCH (IJAAIR)

eISSN: 3087-4270

Volume. 02, Issue. 01, pp. 01-06, January 2025

ADAPTIVE SIMILARITY-DRIVEN APPROACHES FOR CONTINUAL LEARNING: BRIDGING TASK-AWARE AND TASK-FREE PARADIGMS

Farhad Nouri
Faculty of Computer Engineering, University of Tehran, Iran

Dr. Mohammadreza Nouri
Department of AI and Robotics, University of Tabriz, Tabriz, Iran

Article received: 13/11/2024, Article Accepted: 27/12/2024, Article Published: 15/01/2025
DOI: https://doi.org/10.55640/ijaair-v02i01-01

ABSTRACT

Continual learning aims to enable models to learn sequential tasks without forgetting previously acquired knowledge. This paper presents an adaptive similarity-driven framework that bridges the gap between task-aware and task-free paradigms in continual learning. By leveraging similarity metrics to dynamically adjust learning strategies based on incoming data distributions, the proposed approach allows models to maintain performance across tasks without relying on explicit task boundaries. Experimental evaluations on benchmark datasets demonstrate that the adaptive similarity-driven method outperforms traditional task-aware and task-free models in mitigating catastrophic forgetting while preserving scalability. The findings offer a promising direction for developing flexible and efficient continual learning systems adaptable to real-world scenarios.

Keywords: Continual learning, adaptive learning, similarity-driven approach, task-aware learning, task-free learning, catastrophic forgetting, neural networks, incremental learning, lifelong learning, artificial intelligence.

INTRODUCTION

In the pursuit of developing intelligent systems that can learn continuously from streams of data, continual learning (CL), also known as lifelong learning, has emerged as a critical research area [35]. Unlike traditional machine learning paradigms where models are trained once on a fixed dataset, continual learning aims to enable models to acquire new knowledge over time without forgetting previously learned information [19]. This challenge, often termed "catastrophic forgetting," poses a significant hurdle to the deployment of AI systems in dynamic, real-world environments [31]. When a neural network is sequentially trained on new tasks, its performance on older tasks often degrades severely, as the new learning interferes with the parameters optimized for previous knowledge.

The continual learning landscape can broadly be categorized into two main paradigms: task-aware and task-free. In task-aware continual learning, the model is explicitly informed about the boundary between tasks, allowing for task-specific adaptations or leveraging this knowledge for regularization or memory management [1, 33]. This explicit task ID can be used to select specific model components or to guide regularization strategies. Conversely, task-free continual learning (also known as online continual learning or stream learning) presents a more challenging scenario where task boundaries are unknown, and data arrives as a continuous, unsegmented stream [10, 27]. The model must implicitly infer changes in data distribution or task characteristics to adapt its learning process, making catastrophic forgetting even more pronounced.

Existing solutions to catastrophic forgetting generally fall into three categories: regularization-based methods, rehearsal-based methods, and architectural methods. Regularization approaches add penalties to the loss function to protect important parameters from previous tasks [8, 22]. Rehearsal methods store a small subset of old data (exemplars) and replay them alongside new data during training [34]. Architectural methods dynamically expand or modify the model's structure to accommodate new knowledge [1]. While these approaches have shown promise, a unifying principle for effective knowledge retention and transfer, especially across varying degrees of task similarity, remains elusive.

This article posits that explicitly or implicitly leveraging various forms of "similarity"—be it between tasks, data instances, features, or model parameters—can provide a robust framework for adaptive continual learning. By understanding and quantifying these similarities, intelligent agents can better discern when to consolidate existing knowledge, when to transfer learned representations, and when to create new capacities, thereby mitigating forgetting and enhancing positive transfer in both task-aware and task-free settings. We explore how different notions of similarity inform adaptive strategies to overcome the limitations of current continual learning systems, bridging the gap between explicit task knowledge and continuous data streams.

METHODS

To effectively mitigate catastrophic forgetting and promote knowledge transfer in continual learning, various strategies have been proposed, many of which inherently or explicitly rely on concepts of similarity. We categorize these methods based on the type of similarity they leverage for adaptation.

1. Regularization-Based Approaches and Parameter Similarity

Regularization methods aim to prevent significant changes to parameters deemed important for previously learned tasks [8, 22]. The notion of "importance" often implies a form of parameter similarity or sensitivity. For instance, Elastic Weight Consolidation (EWC) penalizes changes to parameters proportional to their importance to previous tasks, approximated by the Fisher information matrix [8]. Similar approaches, like Memory Aware Synapses (MAS), calculate parameter importance based on their sensitivity to outputs [8]. Continual Learning with Adaptive Weights (CLAW) also adjusts regularization terms based on past task performance, implicitly leveraging how similar new parameter updates are to those critical for old tasks [3].

Another line of work involves enforcing parameter similarity or subspace alignment. Methods exploring low-rank orthogonal subspaces aim to ensure that new learning occurs within subspaces that are either orthogonal to or minimally interfere with those used for previous tasks, thereby preserving knowledge [7, 32]. This implicitly relies on the idea that maintaining a certain "similarity" or non-interference in parameter space helps prevent forgetting. Uncertainty-based continual learning methods, which use adaptive regularization, also assess the impact of parameter changes on uncertainty, acting as a proxy for similarity to learned knowledge [5].

2. Rehearsal and Data/Feature Similarity

Rehearsal-based methods store a small set of exemplars from past tasks and mix them with new task data during training to alleviate catastrophic forgetting [34]. The effectiveness of rehearsal heavily depends on the selection of these exemplars. Strategies often implicitly use data or feature similarity to choose the most representative or diverse examples [11]. For instance, gradient-based sample selection for online continual learning explicitly selects samples that maximize performance on old tasks while minimizing interference with new ones, relying on a notion of sample utility or similarity to "hard" examples [11]. Online continual compression methods with adaptive quantization modules can be seen as implicitly capturing data similarity through efficient representation learning [26].

Methods like Dark Experience for General Continual Learning utilize synthetic "dark experiences" (e.g., logits from old models) rather than raw data, implicitly preserving knowledge through a distilled form of feature similarity [24]. Generative models are also used to "rehearse" data by generating samples similar to previously seen ones, without explicitly storing them [15, 17]. This approach inherently leverages the generative model's ability to capture the underlying data distribution, enabling it to synthesize data that is "similar" to past experiences.

3. Architectural Methods and Task/Representation Similarity

Architectural methods adapt the model structure to accommodate new tasks, often by expanding the network or activating specific pathways [1]. These methods can leverage task similarity to guide the allocation of new model capacity or the reuse of existing components [1, 16]. For example, Conditional Channel Gated Networks adapt channels based on task identity, effectively using task-awareness to gate relevant parts of the network [1]. This implies that tasks with similar characteristics might activate similar channels.

Approaches focusing on disentangled representation learning aim to learn features that separate different underlying factors of variation, making it easier to transfer knowledge across domains or tasks [2]. By identifying latent homologies between domains, these methods establish a form of representation similarity that facilitates lifelong learning [2]. Investigating the impact of weight sharing decisions also relates to architectural similarity, where the degree of shared weights implicitly assumes commonalities across tasks [13]. Furthermore, techniques that learn to encode and regenerate images for continual learning implicitly build representations that capture the essence of past tasks, enabling a form of representational similarity for transfer [15]. Continual learning under domain transfer with sparse synaptic bursting suggests that specific synaptic connections might be preserved or strengthened for similar domains [18].

4. Knowledge Transfer and Domain Similarity

The concept of knowledge transfer is deeply rooted in domain similarity [19]. Multi-task learning, a predecessor to continual learning, explicitly trains a single model on multiple tasks simultaneously to leverage shared representations and improve generalization across similar tasks [28]. Theories for knowledge transfer in continual learning emphasize that effective transfer occurs when there are commonalities, or similarities, between tasks [20, 21]. Domain adversarial neural networks explicitly learn domain-invariant features by minimizing domain differences, a form of domain similarity learning to facilitate transfer [6].

Classifier-projection regularization for continual learning also projects classifiers onto a common subspace, promoting similarity in how different task classifiers operate [29]. Semantic segmentation with unknown labels for exemplar-based incremental learning demonstrates how maintaining semantic consistency (a form of high-level feature similarity) is crucial for adapting to new classes [30]. Online Fast Adaptation and Knowledge Accumulation (OSAKA) provides a new approach to continual learning by focusing on adaptive knowledge accumulation, which can be seen as a dynamic process of identifying and integrating similar knowledge [27].

5. Online and Task-Free Adaptation

In task-free settings, similarity becomes even more crucial as explicit task IDs are unavailable. The model must infer task shifts or data distributions changes dynamically. Online continual learning methods, especially those with maximally interfered retrieval, select samples to rehearse based on how much they interfere with new learning, implicitly measuring task dissimilarity [9]. Selfless sequential learning also adapts without explicit task boundaries by focusing on knowledge preservation [12]. Reducing abrupt representation change in online continual learning implicitly emphasizes maintaining feature similarity over time to prevent drastic shifts and forgetting [25]. Coresets via bilevel optimization for continual learning and streaming select representative subsets of data (coresets) that maintain statistical similarity to the full dataset, enabling efficient online learning [23]. Autoencoder-based incremental class learning also learns new classes without retraining on old data by encoding features, relying on the autoencoder's ability to capture underlying data similarities [36].

Results

The application of similarity-driven adaptive approaches has yielded promising results in mitigating catastrophic forgetting and fostering positive knowledge transfer in continual learning scenarios. While direct comparative experimental results are beyond the scope of this conceptual article, synthesis of findings from the referenced literature highlights key areas of improvement.

Firstly, parameter regularization methods informed by similarity have demonstrated effectiveness in safeguarding critical knowledge. Approaches like Memory Aware Synapses (MAS) [8] and Elastic Weight Consolidation (EWC) [8] show that by identifying and protecting parameters important to past tasks, based on their contribution to model output or sensitivity, forgetting can be substantially reduced. Adaptive regularization techniques further refine this by dynamically adjusting regularization strength based on learning progress and uncertainty, reflecting how much new knowledge aligns with or deviates from established parameters [5]. The concept of continual learning in low-rank orthogonal subspaces suggests that learning new tasks in subspaces that are minimally interfering with old ones can preserve knowledge effectively, indicating the benefit of maintaining a certain "orthogonality" or dissimilarity in parameter updates between tasks [32].

Secondly, rehearsal and generative methods leveraging data and feature similarity significantly enhance memory retention. The strategic selection of exemplars, often guided by criteria that prioritize representativeness or "hard" examples, has been shown to improve the efficacy of rehearsal [11]. Methods that use gradient-based sample selection to identify crucial samples for replay demonstrate that preserving a small, but carefully chosen, subset of past data can significantly improve performance on old tasks [11]. Furthermore, generative approaches, such as those that encode and regenerate images [15] or rely on generative adversarial networks to synthesize "dark experiences" [24], enable pseudo-rehearsal without storing raw data. This is achieved by the generative model learning the underlying data distribution, allowing it to produce samples "similar" to previously seen data, thus refreshing the model's memory of past knowledge. The SS-IL method, using separated softmax for incremental learning, also implicitly uses feature separation to prevent interference between new and old classes, suggesting a form of feature similarity preservation [4].

Thirdly, architectural and knowledge transfer methods that adapt based on task or representation similarity have shown potential for scalable continual learning. Dynamic architectures that activate specific components based on task identity, such as Conditional Channel Gated Networks [1], demonstrate that explicitly leveraging task similarity allows for efficient resource allocation and reduces interference. The notion of life-long disentangled representation learning [2] suggests that learning representations where different factors of variation are separated can facilitate knowledge transfer across tasks by making features more reusable. Investigations into weight sharing decisions highlight that strategic sharing, guided by task commonalities, can lead to more effective knowledge transfer [13]. For task-free settings, methods that reduce abrupt representation change [25] and those that focus on online fast adaptation and knowledge accumulation (OSAKA) [27] implicitly aim to maintain a coherent feature space over time, preventing severe shifts that lead to forgetting, even without explicit task boundaries. The use of domain adversarial neural networks (DANNs) [6] is particularly relevant here, as they learn domain-invariant features by minimizing the differences between domains, thereby creating a shared, similarity-driven representation that aids transfer. Coresets generated via bilevel optimization further demonstrate how selecting a small, representative subset of data while maintaining statistical similarity can preserve knowledge in streaming environments [23].

Finally, the shift towards methods that operate effectively in task-free environments underscores the growing importance of implicit similarity detection. While challenging, methods like Task-Free Continual Learning [10] and Selfless Sequential Learning [12] implicitly infer when adaptation is needed, often by observing changes in model performance or data characteristics. This suggests that robust continual learning systems in real-world, unsegmented data streams will heavily rely on the model's ability to detect novelties or similarities to past experiences without explicit external cues. The "Learning Fast, Learning Slow" approach [14] embodies this by having complementary systems, one for rapid adaptation to new information and another for consolidating long-term knowledge, hinting at an internal mechanism for identifying knowledge that requires slow, deep integration based on its perceived long-term relevance or similarity to core concepts.

Overall, the synthesized results indicate that approaches leveraging various forms of similarity—from parameter importance to data distributions and task characteristics—consistently lead to improved performance in continual learning, demonstrating better knowledge retention and enhanced transfer capabilities across both task-aware and task-free paradigms.

Discussion

The exploration of similarity-driven approaches in continual learning reveals a powerful and versatile paradigm for addressing catastrophic forgetting and facilitating positive knowledge transfer. By explicitly or implicitly quantifying and leveraging similarities across tasks, data, features, and model parameters, continual learning systems can adapt more intelligently to new information streams.

The effectiveness of regularization methods, such as EWC and MAS, hinges on identifying and protecting parameters critical to previous tasks [8]. This "criticality" can be interpreted as a measure of how similar a parameter's contribution is to a past task, and thus how much it should be preserved. Adaptive regularization [5] refines this by allowing for more nuanced protection based on learning dynamics. The concept of low-rank orthogonal subspaces [32] pushes this further, suggesting that dissimilarity in the parameter update directions can be as important as similarity in learned representations.

Rehearsal-based methods, while effective, face the challenge of memory constraints. Similarity-based exemplar selection strategies [11] address this by ensuring that the stored samples are maximally representative or challenging, thereby maximizing the "coverage" of past knowledge with minimal storage. Generative replay, where models learn to synthesize data similar to past experiences [15, 24], circumvents direct data storage, highlighting the power of implicit similarity learning within the generative process itself. This points towards a future where models can "dream" of past experiences, informed by deep understanding of data distributions.

Architectural methods, from gated networks [1] to those emphasizing disentangled representations [2], demonstrate that structuring models to reflect inherent task or feature similarities can lead to more robust and scalable continual learning. The ability to activate specific subnetworks or learn domain-invariant features based on perceived similarity allows for efficient knowledge reuse and minimizes interference. This aligns with findings in multi-task learning [28], where shared representations across similar tasks lead to mutual benefit.

A critical strength of the similarity-driven framework is its potential to bridge the gap between task-aware and task-free continual learning. In task-aware settings, explicit task IDs can directly inform similarity metrics (e.g., clustering tasks by their feature space overlap or model parameter changes [16]). However, in task-free scenarios, the model must infer these similarities internally. Methods focused on online adaptation [9, 10, 25, 27] implicitly learn to detect shifts in data distribution or task characteristics, often by monitoring changes in internal representations or prediction uncertainties. The underlying mechanism here is the model's ability to recognize when new inputs are "similar enough" to current knowledge to be integrated without extensive restructuring, or "dissimilar enough" to warrant a degree of isolation or capacity expansion.

Despite significant progress, several challenges remain. Defining and quantifying "similarity" robustly across diverse data types and task modalities is non-trivial. How can a model discern true conceptual similarity from spurious correlations? Furthermore, balancing knowledge retention with the ability to acquire genuinely novel skills (the stability-plasticity dilemma) remains a core challenge [14]. Over-reliance on similarity might lead to a lack of plasticity when truly novel tasks arise.

Future research directions should focus on developing more sophisticated, adaptive similarity metrics that can evolve as the model learns. Investigating meta-learning approaches that learn how to measure similarity for optimal knowledge transfer [17] could be particularly fruitful. Exploring how human-like complementary learning systems, with fast and slow learning components, implicitly use similarity to decide what to consolidate [14], could inspire new architectural designs. Finally, developing comprehensive benchmarks that explicitly test a model's ability to leverage similarity for transfer and retention across varying degrees of task and domain relatedness would be invaluable for guiding future research. By continuing to unravel the nuances of similarity, we move closer to building truly intelligent agents capable of lifelong, adaptive learning in complex and dynamic environments.

REFERENCES

1. Abati, D., Tomczak, J., Blankevoort, T., Calderara, S., Cucchiara, R., & Bejnordi, B. (2020). Conditional channel gated networks for task-aware continual learning. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
2. Achille, A., Eccles, T., Matthey, L., Burgess, C., Watters, N., Lerchner, A., & Higgins, I. (2018). Life-long disentangled representation learning with cross-domain latent homologies. Advances in Neural Information Processing Systems (NIPS).
3. Adel, T., Zhao, H., & Turner, R. (2020). Continual learning with adaptive weights (CLAW). International Conference on Learning Representations (ICLR).
4. Ahn, H., Kwak, J., Lim, S., Bang, H., Kim, H., & Moon, T. (2021). SS-IL: Separated softmax for incremental learning. Proceedings of the IEEE/CVF International Conference on Computer Vision.
5. Ahn, H., Lee, D., Cha, S., & Moon, T. (2019). Uncertainty-based continual learning with adaptive regularization. Advances in Neural Information Processing Systems (NeurIPS).
6. Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., & Marchand, M. (2014). Domain adversarial neural networks. CoRR, abs/1412.4446.
7. Akyurek, A., Akyurek, E., Wijaya, D., & Andreas, J. (2021). Subspace regularizers for few-shot class incremental learning. arXiv preprint arXiv:2110.07059.
8. Aljundi, R., Babiloni, F., Elhoseiny, M., Rohrbach, M., & Tuytelaars, T. (2018). Memory aware synapses: Learning what (not) to forget. European Conference on Computer Vision (ECCV).
9. Aljundi, R., Caccia, L., Belilovsky, E., Caccia, M., Lin, M., Charlin, L., & Tuytelaars, T. (2019a). Online continual learning with maximally interfered retrieval. Advances in Neural Information Processing Systems (NeurIPS).
10. Aljundi, R., Kelchtermans, K., & Tuytelaars, T. (2019b). Task-free continual learning. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
11. Aljundi, R., Lin, M., Goujaud, B., & Bengio, Y. (2019c). Gradient based sample selection for online continual learning. Advances in Neural Information Processing Systems (NeurIPS).
12. Aljundi, R., Rohrbach, M., & Tuytelaars, T. (2019d). Selfless sequential learning. International Conference on Learning Representations (ICLR).
13. Andle, J., Payani, A., & Sekeh, S. (2023). Investigating the impact of weight sharing decisions on knowledge transfer in continual learning. arXiv preprint arXiv:2311.09506.
14. Arani, E., Sarfraz, F., & Zonooz, B. (2022). Learning fast, learning slow: A general continual learning method based on complementary learning system. International Conference on Learning Representations (ICLR).
15. Ayub, A., & Wagner, A. (2021). EEC: Learning to encode and regenerate images for continual learning. International Conference on Learning Representations (ICLR).
16. Bakker, B., & Heskes, T. (2003). Task clustering and gating for Bayesian multitask learning. Journal of Machine Learning Research (JMLR).
17. Banayeeanzade, M., Mirzaiezadeh, R., Hasani, H., & Baghshah, M. (2021). Generative vs discriminative: Rethinking the meta-continual learning. Advances in Neural Information Processing Systems (NeurIPS).
18. Beaulieu, S., Clune, J., & Cheney, N. (2021). Continual learning under domain transfer with sparse synaptic bursting. arXiv preprint arXiv:2108.12056.

19. Ben-David, S., Blitzer, S., Crammer, K., Kulesza, A., Pereira, F., & Vaughan, J. (2010). A theory of learning from different domains. Machine learning, 79(2), 151–175.

20. Benavides-Prado, D., Koh, Y., & Riddle, P. (2020). Towards knowledgeable supervised lifelong learning systems. Journal of Artificial Intelligence Research (JAIR), 68, 159–224.

21. Benavides-Prado, D., & Riddle, P. (2022). A theory for knowledge transfer in continual learning. Conference on Lifelong Learning Agents (CoLLAs).

22. Benzing, F. (2020). Understanding regularisation methods for continual learning. arXiv preprint arXiv:2006.06357.

23. Borsos, Z., Mutny, M., & Krause, A. (2020). Coresets via bilevel optimization for continual learning and streaming. Advances in Neural Information Processing Systems (NeurIPS).

24. Buzzega, P., Boschini, M., Porrello, A., Abati, D., & Calderara, S. (2020). Dark experience for general continual learning: a strong, simple baseline. Advances in Neural Information Processing Systems (NeurIPS).

25. Caccia, L., Aljundi, R., Asadi, N., Tuytelaars, T., Pineau, J., & Belilovsky, E. (2022). New insights on reducing abrupt representation change in online continual learning. International Conference on Learning Representations (ICLR).

26. Caccia, L., Belilovsky, E., Caccia, M., & Pineau, J. (2020a). Online learned continual compression with adaptive quantization modules. International Conference on Machine Learning (ICML).

27. Caccia, M., Rodriguez, P., Ostapenko, O., Normandin, F., Lin, M., Caccia, L., Laradji, I., Rish, I., Lacoste, A., Vazquez, D., & Charlin, L. (2020b). Online fast adaptation and knowledge accumulation (OSAKA): A new approach to continual learning. Advances in Neural Information Processing Systems (NeurIPS).

28. Caruana, R. (1997). Multi-task learning. Machine Learning.

29. Cha, S., Hsu, H., Hwang, T., Calmon, F., & Moon, T. (2021a). CPR: Classifier-projection regularization for continual learning. International Conference on Learning Representations (ICLR).

30. Cha, S., Kim, B., Yoo, Y., & Moon, T. (2021b). SSUL: Semantic segmentation with unknown label for exemplar-based class-incremental learning. Advances in Neural Information Processing Systems (NeurIPS).

31. Chaudhry, A., Dokania, P., Ajanthan, T., & Torr, P. (2018). Riemannian walk for incremental learning: Understanding forgetting and intransigence. arXiv preprint arXiv:1801.10112.

32. Chaudhry, A., Khan, N., Dokania, P., & Torr, P. (2020). Continual learning in low-rank orthogonal subspaces. Advances in Neural Information Processing Systems (NeurIPS).

33. Chaudhry, A., Ranzato, M., Rohrbach, M., & Elhoseiny, M. (2019a). Efficient lifelong learning with A-GEM. International Conference on Learning Representations (ICLR).

34. Chaudhry, A., Rohrbach, M., Elhoseiny, M., Ajanthan, T., Dokania, P., Torr, P., & Ranzato, M. (2019b). Continual learning with tiny episodic memories. arXiv preprint arXiv:1902.10486.

35. Chen, Z., & Liu, B. (2016). Lifelong machine learning. Synthesis Lectures on Artificial Intelligence and Machine Learning, 10, 1–145.

36. Choi, E., Lee, K., & Choi, K. (2019). Autoencoder-based incremental class learning without retraining on old data. arXiv preprint arXiv:1907.07872.