Beyond Recall: Evaluating Forgetting Mechanisms for Multi-Modal Episodic Robotic Memory

Joana Plewnia, Fabian Ternava and Tamim Asfour

***Abstract*— Robot cognitive architectures increasingly emphasize the importance of memory as an active component, bridging the gap between semantic understanding and sensorimotor experiences. It not only manages the flow of information between different processes but also provides essential services: extracting semantic meaning from sensorimotor data, parameterizing symbolic plans with actionable parameters, and predicting the outcomes of actions. Crucially, such memory systems must not only acquire and retain information but also selectively forget - a process that prevents information overload, reduces computational overhead, and helps maintain relevant, up-to-date knowledge for decision-making. This extended abstract describes our work on implementing forgetting mechanisms in the deep episodic memory of a humanoid robot. We report on the results of various experiments to describe the methods used to evaluate the effects of forgetting mechanisms on robot task performance over extended periods in real-world experiments.**

I. Introduction

While recent advances in robotic systems demonstrate promising capabilities in controlled environments, significant challenges remain in developing robots that can operate robustly in dynamic, unstructured settings. The transition from traditional industrial applications to complex domains such as domestic environments, healthcare assistance, autonomous warehouse management, and disaster response requires fundamental advances in adaptive learning capabilities. Central to this development is the implementation of effective episodic memory systems that enable experience-based knowledge acquisition and retrieval. However, comprehensive storage of high-dimensional sensorimotor data presents significant computational and architectural challenges, primarily due to data volume scalability constraints and temporal-spatial redundancies in the acquired information. Drawing parallels from cognitive neuroscience, particularly human memory consolidation processes, selective forgetting mechanisms emerge as a theoretically grounded approach for optimal information retention. These mechanisms could enable discriminative data preservation while systematically eliminating redundant or obsolete information, thereby maintaining computational efficiency and memory relevance.

This work has been supported by the Carl Zeiss Foundation through the JuBot project, by the Baden-Württemberg Ministry of Science, Research and the Arts (MWK) as part of the state's "digital@bw" digitization strategy in the context of the Real-World Lab "Robotics AI.", by the European Union's Horizon Europe Framework Programme under grant agreement No 101070596 (euROBIN), and by the German Federal Ministry of Education and Research (BMBF) under the Robotics Institute Germany (RIG).

The authors are with the High Performance Humanoid Technologies Lab, Institute for Anthropomatics and Robotics, Karlsruhe Institute of Technology (KIT), Germany. E-mails: {joana.plewnia,asfour}@kit.edu.

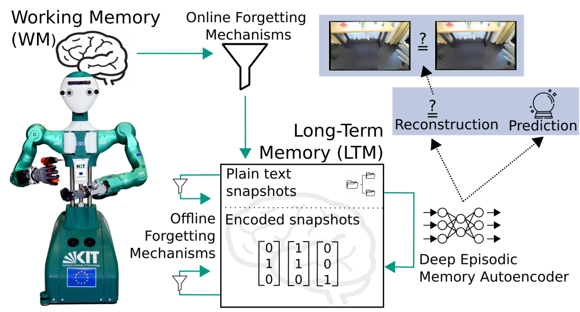

Fig. 1. We evaluate the performance of different forgetting mechanisms from [1] applied to the humanoid robot ARMAR-6 [3] and the *20 Billion Something-Something v2* dataset [4], based on the reconstruction and prediction capabilities of the Deep Episodic Memory Autoencoder trained on the resulting subset of memory snapshots. Reconstruction and prediction capabilities are analyzed by calculating PSNR and SSIM values compared to the ground truth memory snapshot.

This extended abstract describes the extension of our work on forgetting mechanisms in episodic memory [1] of the cognitive architecture of the ARMAR humanoid robots [2], by extending the evaluation to a large real-world dataset collected from the ARMAR-6 humanoid robot [3]. We discuss the impact of forgetting mechanisms on long-term episodic memory performance over extended periods, incorporating real-world sensory inputs and additional benchmark datasets. The findings aim to provide insights into designing efficient and adaptive memory architectures for real-world robotic applications, advancing their adaptability and efficiency in complex, dynamic tasks.

II. Approach

The cognitive robot architecture of our ARMAR humanoid robots, which is implemented in ArmarX[1], differentiates four types of memory: (i) Sensory Memory (SM), which temporarily stores raw sensor input before transfer to the working memory, (ii) Working Memory (WM), which maintains the robot's current internal and external state representation, (iii) Long-Term Memory (LTM), which stores knowledge no longer actively needed in the working memory, and (iv) Prior Knowledge (PK), which contains developer-defined knowledge loaded into the working memory during initialization. This structure facilitates information flow from perception through processing to long-term storage. Our long-term memory consists of two main elements: (i) a memory cache storing plain-text snapshots on a hard drive and

[1]https://armarx.humanoids.kit.edu/

(ii) a deep episodic memory that compresses snapshots into latent space representations using Auto-Encoders (AE). For a more detailed description of the robot cognitive architecture and the implemented memory system, the reader is referred to [2].

Our robot cognitive architecture in ArmarX implements *two* types of forgetting mechanisms: *online and offline forgetting*.

Online forgetting occurs during/before the transfer of entity snapshots from working memory (WM) to long-term memory (LTM). Each snapshot is evaluated based on consolidation frequency and similarity to previously stored data. Only those surpassing predefined relevance and distinctiveness thresholds are retained, while redundant or insignificant snapshots are discarded.

Offline forgetting periodically prunes outdated data from the LTM cache through time-based decay. The remaining snapshots are transformed into latent space representations within the deep episodic memory, where further consolidation occurs based on similarity measures. This ensures that only the most salient information is preserved, while obsolete data is permanently removed.

To refine the memory representations, the deep episodic memory Auto-Encoder (AE) is periodically re-trained before each new forgetting cycle. For a detailed description of the forgetting mechanisms and their implementation in ArmarX, we refer to [1].

We evaluate the impact of forgetting mechanisms on deep episodic memory by analyzing *compression efficiency, reconstruction accuracy*, and *prediction performance*.

For *reconstruction accuracy*, snapshots are encoded and decoded by the autoencoder (AE) of the deep episodic memory. The reconstructed outputs are compared to the originals using Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM), as they are typical metrics for evaluating reconstruction tasks. In the prediction task, the AE receives a snapshot and time-delta to generate a future state, and the prediction performance is evaluated using the same metrics. We investigate different forgetting strategies, including frequency-based, similarity-based and random-based online forgetting methods, as well as time-based decay and offline similarity-based methods.

To quantify the effect of forgetting, we compare deep episodic memories trained on recorded data without forgetting to those using different strategies proposed by [1]. Both quantitative (PSNR, SSIM) and qualitative analyses are conducted to evaluate the impact on reconstruction and prediction. The evaluation is performed on visual data from the *20 Billion Something-Something-v2 (20BN)* dataset [4], as well as multi-modal data recorded with the ARMAR-6 robot, which includes vision, laser scanner, and proprioception modalities. The ARMAR-6 robot operated in a controlled model warehouse environment, executing humanoid-human collaborative tasks such as pick-and-place, and object handovers tasks.

III. Results

Our evaluation shows that online frequency- and similarity-based forgetting mechanisms applied to the 20BN dataset reduced memory size by 80% while maintaining minimal performance degradation in reconstruction and prediction. Similarly, random-based forgetting mechanisms yielded comparable results. In contrast, offline forgetting mechanisms as described in [1], proved ineffective for this dataset, frequently leading to complete data loss. This is attributed to the dataset's short and highly distinct episodes, which hinder the effectiveness of time-based decay and similarity-based consolidation strategies.

For the ARMAR-6 robot's episodic memory, a combination of online frequency-based and similarity-based forgetting mechanisms showed promising results. These strategies reduce the memory size by up to 99% without significant degradation in reconstruction and prediction tasks, as measured by our evaluation metrics PSNR and SSIM. In contrast, random-based forgetting mechanisms resulted in performance degradation in both PSNR and SSIM for smaller reductions. While similarity-based mechanisms outperformed frequency-based ones at higher compression levels, they introduced artifacts when training data was insufficient. Offline similarity-based mechanisms improved reconstruction performance while reducing memory size by 75–84%. However, qualitative analysis revealed limitations in prediction tasks, where the autoencoder (AE) primarily learned to reconstruct inputs rather than predict future states.

IV. Conclusion and Future Work

Online forgetting mechanisms effectively reduce bandwidth and storage requirements for episodic memory, achieving up to 99% size reduction without compromising reconstruction or prediction performance. These mechanisms are particularly suited for dynamic, real-time environments, as they operate during the transfer of snapshot to long-term memory and thus reduce bandwidth usage between temporary and persistent storage. Offline similarity-based strategies enhance reconstruction performance but fail to improve prediction capabilities for the evaluated model. In real-world robotic experiments, forgetting mechanisms proved valuable in dealing with the vast amount of episodic data: Offline forgetting mechanisms improve the reconstruction capabilities of our deep episodic memory, while online forgetting mechanisms allow to reduce bandwidth demands without negatively affecting reconstruction or prediction capabilities.

Current evaluation metrics like PSNR and SSIM are insufficient for assessing the prediction capabilities of autoencoders, as they do not account for whether the predicted changes accurately reflect the input dynamics. Future work will explore more complex metrics that emphasize content similarity, such as FSIM [5] or LPIPS [6], to improve the assessment of the episodic memory prediction models and enable more meaningful evaluations of forgetting mechanisms in the context of prediction tasks. Additionally, further studies should investigate the generalizability of these findings across different robotic platforms and datasets.

References

[1] Joana Plewnia, Fabian Peller-Konrad, and Tamim Asfour. "Forgetting in Episodic Long-Term Memory of Humanoid Robots". In: *IEEE International Conference on Robotics and Automation (ICRA)*. 2024, pp. 6711–6717.

[2] Fabian Peller-Konrad et al. "A memory system of a robot cognitive architecture and its implementation in ArmarX". In: *Robotics and Autonomous Systems* 164 (2023), pp. 1–20.

[3] Tamim Asfour et al. "ARMAR-6: A High-Performance Humanoid for Human-Robot Collaboration in Real World Scenarios". In: *IEEE Robotics & Automation Magazine* 26.4 (2019), pp. 108–121.

[4] Raghav Goyal et al. *The "something something" video database for learning and evaluating visual common sense*. 2017.

[5] Lin Zhang et al. "FSIM: A Feature Similarity Index for Image Quality Assessment". In: *IEEE Transactions on Image Processing* 20.8 (2011), pp. 2378–2386.

[6] Richard Zhang et al. "The Unreasonable Effectiveness of Deep Features as a Perceptual Metric". In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2018, pp. 586–595.