2024 IEEE International Conference on Robotics and Automation (ICRA 2024)
May 13-17, 2024. Yokohama, Japan

Forgetting in Robotic Episodic Long-Term Memory

Joana Plewnia*, Fabian Peller-Konrad* and Tamim Asfour

Abstract— **Artificial cognitive architectures traditionally rely on complex memory models to encode, store, and retrieve information. However, the conventional practice of transferring all data from working memory (WM) to long-term memory (LTM) leads to high data volumes and challenges in efficient information processing and access. Deciding what information to retain or discard within a robot's LTM is particularly challenging since knowledge about future data utilization is absent. Drawing inspiration from human forgetting this paper implements and evaluates novel forgetting techniques that allow consolidation in the robot's LTM only when new information is encountered. The proposed approach combines fast filtering during data transfer to the robot's LTM with slower yet more precise forgetting mechanisms that are periodically evaluated for offline data deletion inside the LTM. We compare different mechanisms, utilizing metrics such as data similarity, data age, and consolidation frequency. The efficacy of forgetting techniques is evaluated by comparing their performance in a task where two ARMAR robots search through their LTM for past object locations in episodic ego-centric images and robot state data. Experimental results show that our forgetting techniques significantly reduce the space requirements of a robot's LTM while maintaining its capacity to successfully perform tasks relying on LTM information. Notably, similarity-based forgetting methods outperform frequency- and time-based approaches. The combination of online frequency-based, online similarity-based, offline similarity-based, and time-based decay methods shows superior performance compared to using individual forgetting strategies.**

I. Introduction

Human cognition is intrinsically tied to sensory perception and memory formation. However, attempting to consolidate *all* information into persistent memories would quickly result in information overload [1]. Human memory leverages many different complex neurological processing mechanisms to avoid this. Even though there are many differences between human and robotic memory, inspiration can be drawn from those mechanisms to deal with similar issues in robotics [2]. One of those problems arises from robots experiencing their surroundings through sensors that generate a high volume of data in a short amount of time. For example, the humanoid robot ARMAR-6 [3] produces multiple hundreds of gigabytes per hour. Attempting to store all data leads to several problems, including constraints on storage capacity, retrieval speed, and bandwidth limitations during data transmission.

* The authors contributed equally to this paper.

The research leading to these results has received funding from 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" and by the Carl Zeiss Foundation through the JuBot project

The authors are with the High Performance Humanoid Technologies Lab, Institute for Anthropomatics and Robotics, Karlsruhe Institute of Technology (KIT), Germany

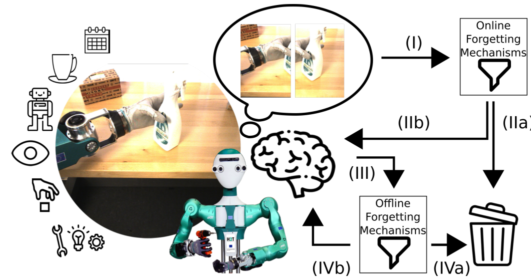

Fig. 1. The humanoid robot ARMAR-6 [3] experiencing a scene and consolidating multi-modal memory snapshots into its LTM. During consolidation online forgetting mechanisms (I) are used to decide if a memory snapshot is directly forgotten (IIa) or consolidated into LTM (IIb). When the robot is no longer in operation offline forgetting mechanisms can be used (III) to further decide on forgetting (IVa) or retaining (IVb) of those memory snapshots.

Storing all information may also harm the performance of learning algorithms that are built in the robot's memory [2]. While there may be a lot of *memory snapshots* to learn from, i. e., information from *entities* recorded or created at a specific point in time, those might not be balanced in a way that is conducive for learning as robots often operate in similar environments, observe similar scenes repeatedly, and frequently encounter comparable data.

Determining which memory snapshots to store is challenging as one would need information about the future to predict what kind of information and thus which memory snapshots will be needed again. The same is true for humans but humans have developed the ability to selectively forget, addressing the issue of retaining excessive amounts of memory snapshots and deciding which ones to consolidate or discard [1].

In this work, we present a human-inspired forgetting model as an integral part of our cognitive robot architecture [4] implemented in the robot software framework ArmarX [5]. Drawing inspiration from human memory processes, our forgetting model integrates two key elements: *Time-based decay* and *retroactive interference* forgetting mechanisms. Time-based decay models human memory decay, while retroactive interference models similarity/cue-overload and mental exertion effects in humans. To the best of our knowledge, time-based and pure displacement mechanisms are more frequently encountered than alternative mechanisms. In contrast, our model uses similarity-based methods in combination with time-based decay and supports online and offline filtering techniques. While many other artificial cognitive architectures do not make use of forgetting in episodic long-term memory [6], our model targets episodic memories specifically as the episodic memory inherently receives new snapshots during the robot's full runtime.

We show that, through a combination of online and offline interference-based and time-based decay mechanisms, our proposed forgetting model decreases the number of memory snapshots consolidated into the robot's multi-modal episodic LTM (as depicted in Fig. 1) while still achieving high performance in an object localization task. Furthermore, we evaluate additional time requirements and space consumption of the episodic memory.

Our findings demonstrate that offline similarity-based forgetting mechanisms achieve the best results if only one single mechanism is used, while the combined use of online frequency- and similarity-based mechanisms, offline time-based decay, and offline similarity-based mechanisms using latent space representations of memory snapshots showed superior overall performance.

II. Related Work

In artificial cognitive systems methods used to forget memories can broadly be categorized based on their objectives and approaches. Some cognitive architectures focus on emulating human behavior, such as forgetting, staying as close as possible to the current theories in psychology [6]. Conversely, others focus on enhancing performance in tasks that rely on working and/or long-term memory using methods that are not directly aligned with human forgetting [7]. Beyond this categorization, artificial forgetting can be further classified by their approach, i. e., whether they employ time-based decay or interference-based mechanisms [8]. While the base concepts of those two mechanisms may be inspired by human forgetting, more complex concepts like emotions have been relatively underexplored [9].

A. Time-based Decay

Being the most prevalent mechanism of forgetting in artificial cognitive systems, time-based decay calculates an initial *activation value* that can be considered as a memory snapshot *strength*, which will then decrease over time by a factor calculated using a decay function. This activation value is then compared to a preset threshold to determine whether the snapshot should be forgotten. This approach mirrors observations in human forgetting, where the strength (and precision) of a memory snapshot seems to decrease with time up to the point where it can't be recalled anymore. Some cognitive architectures calculate this activation value periodically and delete a snapshot if it falls beneath a specific threshold [10], while others calculate this strength only when trying to access a snapshot to determine whether the access operation was successful or not [2].

Pure time-based decay solely considers the amount of time a snapshot has already spent inside the memory when calculating activation values. On creation, each snapshot gets assigned the same activation value and a decreasing forgetting function is used to calculate the activation at a given point in time. Commonly employed decay functions include linear [11] and exponential functions [9], [12], the latter usually approximating the original *Ebbinghaus' curves* derived from human forgetting experiments [13]. Time-based decay does not depend on the modality of memory elements and can be closely adapted from human memory. Thus, time-based decay is often used in robotic research [2], [6], [9], [10], [11], [12], [14], [15]. Prominent examples that allow the use of pure time-based decay are ACT-R [16], employing it in its declarative long-term memory [6] and Soar [17], [18], extending its usage to working memory, procedural memory and declarative long-term memory [6].

Instead of using the same initial activation value for every snapshot, some architectures use importance factors to calculate the initial activation [2], [10], [15]. These importance-factors depend on various values, such as rewards [12], and belief states [2] but also rarely emotions, since research into human forgetting processes has shown that memories associated with strong emotions stay present in the long-term memory for a longer period of time [9], [11], [15], [19]. Such emotions are either incorporated into importance factors [15], used as a factor for the decay rate [9], [19], or as a summand that is added to the activation function, as in Juvina et al. [11]. They successfully demonstrate the incorporation of emotions into ACT-R by introducing an additional module calculating valuation and incorporating it into the activation function. Similarly, LIDA [19] modifies the *reinforcement rate* in response to emotional factors, influencing memory snapshot decay.

B. Interference-based Mechanisms

Interference-based mechanisms operate under the premise that memory space is finite, an observation applicable to both human working memory, which accommodates a limited number of information chunks at any given time [20], and robotic memory constrained by disk space. In robotics, two principal interference-based mechanisms, i. e., *pure displacement*, and *retroactive interference*, are prevalent [8].

Pure displacement replaces existing memory information with newly acquired data, employing criteria such as random selection, recency, or frequency of use. Inspired by established caching strategies in computer science, techniques like least-recently-used (LRU) or first-in-first-out (FIFO) can model pure displacement [8]. As an example, ACT-R employs pure displacement, overwriting older information with new data due to its fixed-size working memory [8].

Retroactive interference models the *cue-overload/similarity* and *mental exertion* effects in human memory. Cue-overload/similarity effects allow humans to only consolidate information if it is considered new and possibly overwrite older similar snapshots associated with the same cue. Mental exertion reflects the cognitive load's impact on memory, resulting in faster forgetting when the mental load is high during initial consolidation [1]. In artificial cognitive architectures, these effects can be modeled by *discriminability values* for memory snapshots as part of the activation value and a mental exertion scaling factor for the initial activation value upon original consolidation into LTM [2].

Interference-based mechanisms can be used in combination with time-based decay. ACT-R, for instance, employs pure displacement in working memory alongside time-based decay in long-term memory. Freedman et al. [2] combine both methodologies into a unified activation function.

Although we believe that forgetting is crucial for cognition, only very few cognitive architectures explicitly include forgetting mechanisms in their memory formulation. For simplicity, it is often assumed to be persistent, because most approaches focus on how the content of memory can be learned effectively [21]. For instance, CRAM [22] does not implement specific forgetting mechanisms, while EPIC explicitly assumes an unlimited working memory [23].

C. Research Gap and Contributions

Even though it is believed that both time-based decay and interference-based mechanisms likely coexist in human forgetting [1], Ricker et al. [24] introduced that these two forgetting mechanisms may be used in different capacities for distinct memory representations. Nonetheless to the best of our knowledge, current literature focuses on time-based decay, as interference-based mechanisms are only found in a limited number of works ([2], [8]).

This work aims to bridge the gap between prevailing forgetting mechanisms in robotic memory and the believed mechanisms of human memory forgetting. We argue that artificial cognitive architectures that try to emulate the human forgetting processes are not complete if they do not consider retroactive interference. Nevertheless, we also argue that forgetting must have a positive effect on task performance. As such, our proposed approach can be placed between approaches emulating human forgetting staying as close as possible to the current theories in psychology and approaches only focussing on task performance.

The forgetting model for our cognitive robot control architecture uses a combination of interference-based and time-based decay mechanisms for multi-modal episodic memory. It takes inspiration from Freedman et al. [2], who uses a combined forgetting approach that only acts as a filter instead of forgetting snapshots permanently.

We implement the proposed forgetting model as part of our cognitive robot control architecture [4] in the robot software framework ArmarX [5] and evaluate different combinations of forgetting mechanisms in the robot's episodic memory using object localization accuracy based on the information stored in the robot's LTM as a metric. In addition, we evaluate the time requirements and space consumption reduction.

III. Proposed Forgetting Model

As depicted in Fig. 2 and Fig. 3, the proposed forgetting model is part of our cognitive robot control architecture [4] in the robot software framework ArmarX [5]. The robot's memory is divided into working and long-term memory. Both WM and LTM consist of distributed memory servers for distinct modalities, like vision, robot state, and object information. Similar to human memory, the working memory of our architecture can only hold a certain amount of data, i. e., entity snapshots such as images captured by the RGB camera of the robot or proprioception information.

Whenever the capacity in the WM for an entity is exceeded, the oldest entity snapshot is transferred into the LTM. Pre-defined filters control whether the entity snapshot will be consolidated into LTM or not. Examples of such filters include frequency-based and similarity-based ones. Frequency-based filters only consolidate an entity snapshot if a certain amount of time δ_{wait} has passed since the last snapshot of this entity was consolidated, mimicking mental exertion in humans. Thus, $\delta_{wait} \leq t(x_i) - t_{now}$. While the parameter δ_{wait} is adjustable, $t(x_i)$ describes the last time a snapshot $x_i \in \mathcal{X}$ of the entity $\mathcal{X}$ was consolidated into the LTM. t_{now} describes the current time. Frequency-based filters work independently of the entity snapshot's modality.

Similarity-based filters only consolidate an entity snapshot if it surpasses a certain amount of dissimilarity regarding the latest n entity snapshots of the same entity $\mathcal{X}$. Thus, an entity snapshot $x_j \in \mathcal{X}$ will only be stored in the LTM if

$$d_{min} \leq \frac{1}{n} \cdot \sum_{i \in \{1..n\}} \mathrm{D}(x_j, x_i) \tag{1}$$

with D being a dissimilarity measure and d_{min} being the minimum dissimilarity an entity snapshot needs to have in comparison to the last n accepted snapshots to be consolidated. d_{min} and n are adjustable parameters. Due to the fact that we need real-time capable algorithms, that decide which snapshots to consolidate from WM to LTM as shown in Fig. 2, we implement mean-squared error (MSE), mean absolute error (MAE), and cosine distance and compare those different dissimilarity metrics regarding their performance.

Additionally, we use modality-independent offline mechanisms that filter already consolidated entity snapshots inside the episodic memory and mimic the decay of memory snapshots over time in humans. A schematic of this part can be seen in Fig. 3. We calculate temporal activation $a_{temporal}(x_i)$ using an exponential function with a negative exponent (Eq. 2) or the Ebbinghaus curve [13].

$$a_{temporal}(x_i) = e^{-\Delta t} \tag{2}$$

where $\Delta t = t_{start} - t(x_i)$ and t_{start} references the time this entity $\mathcal{X}$ was experienced for the first time. The temporal activation is evaluated against a minimal activation value a_{min} and an entity snapshot will be forgotten if its activation falls below this threshold. After this filtering step, the remaining snapshots will be encoded using a previously trained *Deep Episodic Memory* autoencoder [25], [26]. Based on the principle of Wasserstein Autoencoders [27], the learned latent vector representation supports the prediction of entity snapshots, distance measuring, and data compression. Instead of storing a timeline of raw data, we only keep a timeline of latent vectors as well as the encoder and decoder weights. That way, we can extend our deep episodic memory once new knowledge enters the LTM and we can send the stored knowledge back to the WM if recalled by reconstructing it to its original form. The autoencoder is trained in an offline phase. While the filter steps follow an all-or-nothing

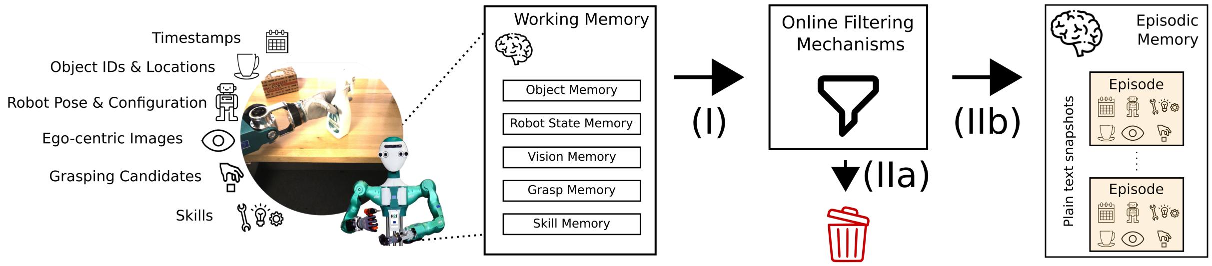

Fig. 2. Structure of our proposed forgetting model for online filtering of entity snapshots upon consolidation from WM to LTM. When the maximum capacity of entity snapshots of this entity in the WM is reached, consolidation into the LTM is considered (I). Depending on pre-defined parameters either frequency- or similarity-based filters are used to determine whether to forget (IIa) or consolidate (IIb) the snapshot into the episodic memory of the robot.

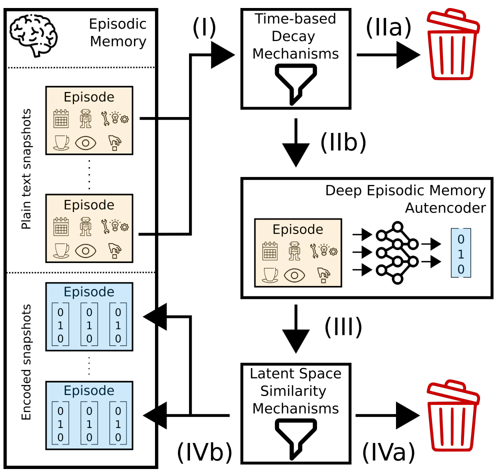

Fig. 3. Structure of our proposed forgetting model for offline filtering of entity snapshots in the episodic memory. After consolidation, information is temporarily stored in plain text in the LTM of the robot (orange). Once the robot stops operation, we employ time-based decay mechanisms (I) to filter out data, that is assumed to be not relevant anymore (IIa). The remaining data (IIb) is converted into a latent representation (blue). Finally, similarity-based filters are used (III). If accepted (IVb), the set of latent vectors is stored in the encoded version of the episodic memory. All entity snapshots that were not accepted are permanently deleted (IIa and IVa). The proposed mechanisms are independent of the memory snapshot's modality.

approach, this encoding procedure is able to filter out irrelevant information within one snapshot, successively removing unnecessary details from the entity, thus "blurrying" the raw input.

Finally, an additional similarity-based latent space filter (Eq. 3) is used, which leverages cosine distance to decide between further consolidation and forgetting. Let $\mathrm{enc}(x_i)$ denote the latent vector representation of entity snapshot x_i. We calculate

$$a_{latent}(x_i) = \frac{1}{n} \cdot \sum_{j \in \{1..n\}} \mathrm{D_C}(\mathrm{enc}(x_i), \mathrm{enc}(x_j)) \quad (3)$$

If $a_{min,latent} \leq a_{latent}(x_i)$, i. e., if it is greater than a minimum required activation for latent activation values, the episode snapshot is kept inside the LTM, if not it is forgotten and removed from memory. Here, $\mathrm{D_C}(x, y)$ denotes cosine distance. Only those snapshots, that pass these filters are stored or kept in the episodic memory. Fig. 2 shows structurally, how and when we employ a forgetting mechanism to filter out snapshots before and after encoding.

This proposed structure allows combining both, time-based decay and similarity-based forgetting mechanisms, as well as forgetting memory snapshots permanently while leveraging a combination of online and offline mechanisms.

IV. Experimental Setup

To test our proposed forgetting model, we designed an experiment focused on a task that relies on the robot's memory. For that, we recorded the episodic memory of the two humanoid robots ARMAR-III (simulation) and ARMAR-6 (real robot experiments) during typical household tasks like picking up an object and bringing it to a different location or handing it over to the human using different filter methods and parameterizations. However, our model is not limited to humanoid robots.

One fundamental challenge faced by both humans and robots involves recollecting the precise location where they last encountered an object. During our experiments, we placed multiple objects within the robot's field of view, even though the robot didn't actively interact with these objects during its assigned task. In our simulated robot experiments we placed 40 different objects in the simulated room, while we placed 23 objects in the room during our real robot experiments. To stay close to an intuitive human approach when trying to remember where they last saw an object, we evaluate our proposed model by searching through the episodic memory. For our evaluation, we analyzed the consolidated entity snapshots of the robot's visual perception stored in the episodic memory by searching from the most current entity snapshot backward in time until we found the object. Then we queried the robot state from the episodic memory, i. e., the robot's configuration and position in global coordinates, to find out where the robot was at the time the entity snapshot was recorded. Both, visual information and robot state information have been filtered using different forgetting mechanisms. In our experiments we focussed on the most used, most frequent, and most storage-intensive data sources. Thus, the used modalities should be seen as an example since our memory and forgetting model can easily be extended to other modalities.

Throughout the experiments, we quantified the following metrics: (1) Object localization accuracy: We collected how many objects the robot accurately recognized in the given task, comparing this with performance when no forgetting mechanisms were used. (2) Reduction in episodic memory size: We evaluated to which extent the memory size decreased when incorporating different forgetting mechanisms.

For comparability of the different mechanisms (online, and offline before or after latent compression), we only take the amount of snapshots into account. (3) Time requirements: We measured the overhead associated with online forgetting (during the task) and offline forgetting (when the robot is no longer in operation). All experiments were performed on a computer with an AMD Ryzen 9 3900X 12-Core Processor, an NVIDIA GeForce RTX 3060, and 80 GB RAM. We systematically compared various forgetting mechanisms and combinations thereof using identical episodes, subsequently reporting the parameter settings that yielded the best outcomes.

V. Results

As outlined in Section III, we implement and assess two categories of forgetting mechanisms using two different modalities as an example: Frequency-based mechanisms and similarity-based mechanisms being applied to image and proprioception data. In Table I, we present the most significant size reductions achieved by each forgetting mechanism and their combinations that still yielded 100% correct object location detections.

A. Online Forgetting

The first five entries of Table I display results for individual mechanisms, including frequency-based, similarity-based with MSE, and cosine similarity, time-based decay, and similarity-based forgetting utilizing latent space representations. Rows 1–3 represent online consolidation of WM to LTM, while rows 4–5 illustrate offline consolidation. Additionally, we evaluated combinations of the proposed mechanisms (rows 6–16). As MSE and MAE delivered comparable results in our experiment, both were summarised under MSE.

Frequency-based mechanisms (1) can reduce stored snapshots by up to 84.72%. However, similarity-based ones (2–3) surpass frequency-based mechanisms slightly, enabling LTM size reduction by up to 87.69%. The reason for that is that robots often operate in similar environments, observe similar scenes repeatedly, and frequently encounter comparable data. Thus, similarity-based mechanisms are better suited to prune unnecessary information and allow for a greater reduction in LTM size without negatively affecting task performance, even if simple (in terms of implementation) similarity measures like MSE are used. In contrast, frequency-based approaches are way easier to compute. Table I also shows that similarity-based mechanisms in general, but complex similarity measures like cosine similarity in particular have a greater cost regarding the additional computation time during the forgetting process although there are only small differences in performance. The best results are achieved by MSE (2). Notably, employing complex similarity measures like cosine similarity implies additional computational costs during the forgetting process. For instance, in our experiments, applying cosine similarity led to a time increase of 57.06%, whereas frequency-based mechanisms added less than 0.01% to computation time.

Increasing the consolidation threshold parameter d_{min}, as illustrated in Fig. 4, results in a reduction of the number of correctly localized objects. This leads to instances where some objects are detected at incorrect locations, while others become entirely undetectable. If the threshold is further increased, the number of undetectable elements continues to rise until no objects can be detected at all.

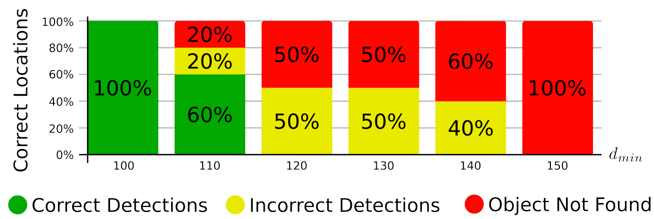

Fig. 4. Similarity-based online forgetting using MSE showing the amount of correct and incorrect object detections and objects that were not found, compared for different values of d_{min}.

B. Offline Forgetting

Offline forgetting mechanisms were examined, including cosine similarity (Eq. 3) applied exclusively to latent representations of entity snapshots, and time-based decay employing an exponential forgetting function (Eq. 2) or the Ebbinghaus forgetting function.

During our experiments, we achieved additional size reductions up to 55.40% using time-based decay (see Table I (4)) while not negatively affecting task performance. We omitted time-based decay using the Ebbinghaus forgetting function as the results were identical to those of the exponential forgetting function.

Calculating a_{latent} (see Eq. 3) of the latent space representations of an entity for filtering proved to be more effective. We report the highest achievable size reduction of 95.38% for a single forgetting mechanism that still achieved 100% accurate object localizations in Table I (5).

C. Combination of Forgetting Mechanisms

The strength of our proposed forgetting model lies in the flexibility to combine different forgetting mechanisms as shown in Table I (6–16). For instance, leveraging frequency-based and similarity-based online mechanisms as well as time-based decay achieves a higher size reduction than any single forgetting mechanism by 0.86%.

Our results show, that using latent space similarity mechanisms in combination with any other mechanism always improves the achieved size reduction. On the other hand, only online similarity mechanisms should be used in combination with the latent space similarity mechanism (10,13,14). Frequency-based methods can be combined with time-based decay beforehand (9, 11, 16) to improve size reductions.

The most impressive outcomes are achieved by combining all mechanisms (16). This combination reduces the size of the episodic memory to 3.03% of its original size. During our real robot experiments, a combination of frequency-based, online similarity-based (MSE) and offline similarity-based (cosine) mechanisms yielded the best reduction of memory size (up to 7.99% of the original size) while achieving 100% in the object recognition task.

TABLE I

Evaluation of Experiments in Simulation

	Frequency	Similarity			Time-based decay		latent space similarity		additional time		size	
	δ_{wait}	D	d_{min}	n	type	a_{min}	D_C	$a_{min,latent}$	online	offline	best result	on average
(1)	200 ms	-	-	-	-	-	-	-	<0.01 sec	0 sec	15.28 %	15.38 %
(2)	-	MSE	100.0	2	-	-	-	-	8.2 sec	0 sec	12.31 %	12.43 %
(3)	-	Cosine	0.1	2	-	-	-	-	287 sec	0 sec	14.10 %	14.42 %
(4)	-	-	-	-	Exp.	1.679	-	-	0 sec	382 sec	46.60 %	50.77 %
(5)	-	-	-	-	-	-	Cosine	0.3	0 sec	578 sec	**4.62 %**	4.75 %
(6)	200 ms	MSE	50.0	-	-	-	-	-	8.3 sec	0 sec	7.75 %	7.82 %
(7)	200 ms	-	-	-	Exp.	1.679	-	-	<0.01 sec	383 sec	7.12 %	7.48 %
(8)	200 ms	-	-	-	-	-	Cosine	0.1	<0.01 sec	579 sec	6.31 %	6.37 %
(9)	200 ms	MSE	50.0	-	Exp.	1.679	-	-	8.3 sec	402 sec	3.76 %	3.98 %
(10)	200 ms	MSE	50.0	-	-	-	Cosine	0.1	8.3 sec	384 sec	3.14 %	3.21 %
(11)	200 ms	-	-	-	Exp.	1.679	Cosine	0.1	<0.01 sec	962 sec	4.50 %	4.63 %
(12)	-	MSE	100.0	2	Exp.	1.679	-	-	8.2 sec	380 sec	5.72 %	6.04 %
(13)	-	MSE	100.0	2	-	-	Cosine	0.1	8.1 sec	580 sec	3.25 %	3.31 %
(14)	-	MSE	100.0	2	Exp.	1.679	Cosine	0.1	8.0 sec	800 sec	3.07 %	3.14 %
(15)	-	-	-	-	Exp.	1.679	Cosine	0.1	0 sec	961 sec	6.95 %	7.17 %
(16)	200 ms	MSE	100.0	2	Exp.	1.679	Cosine	0.1	10.2 sec	421 sec	**3.03 %**	3.09 %

Best size reductions while maintaining 100% correct object localizations for each mechanism (rows 1–5) and combination of mechanisms (6–16). All mechanisms were used on the same episode. Additionally, we report the average size of episodic memories constructed from $n = 3$ episodes. The episodes have an average duration of 8:32 minutes.

VI. Conclusion

In this work, we successfully implemented and evaluated a novel forgetting model within our cognitive robot architecture in the robot software framework ArmarX using a combination of online frequency- and similarity-based mechanisms and offline time-based decay and similarity-based mechanisms on multi-modal data. Our approach demonstrated its applicability by achieving substantial reductions (up to 96.97%) in the size of the episodic memory without negatively impacting the performance of an object localization task that is fully dependent on the LTM.

When comparing single forgetting mechanisms, offline similarity-based mechanisms leveraging latent space representations showed better performance than frequency-based mechanisms, offline time-based decay or online similarity-based mechanisms. We achieved the most promising results when combining online frequency- and similarity-based mechanisms with offline time-based decay and offline similarity-based mechanisms while incorporating latent space representations of memory snapshots. Using this combination we reduced the size of the episodic memory by 96.97% while needing less than 2% of additional time during consolidation from WM into LTM and less than 200% of additional time when using offline forgetting.

In summary, our research demonstrates that forgetting models are valuable in the context of cognitive robotics, where robots have to deal with large amounts of information. If parameterized and combined correctly, different forgetting mechanisms can have a great influence on episodic memory size while not negatively affecting the task performance of robots. In contrast, a smaller memory size may even support task performance, e. g., through shorter retrieval times.

VII. Discussion and Future work

While we showed that our methods of forgetting are beneficial for reducing the size of episodic memory while not negatively impacting the task of object localization, the impact of our approach on more complex and varied tasks with more modalities should be investigated. As a first step towards this topic, we focused in our experiments on the largest, most used, and most frequent data sources of our robots: Images and data from the robot's proprioception. As of now, these two modalities alone are responsible for more than 90% of the data in the robot's memory. An analysis of which methods should be used for which modality would be desirable. Additionally, further investigation into time-based decay is needed when handling episodes that were recorded over a longer period of time, such as days (or even a lifetime). Since we handcrafted most of the parameters used in this paper, another open question is how one can automatically derive parameters for the proposed mechanisms that work well on unseen, heterogeneous episodes, as we found that the same parameter might lead to very different performances of the forgetting mechanisms if the scenarios differ significantly from each other.

It's important to note that human forgetting does not follow an all-or-nothing approach. Instead, humans tend to forget certain details of an episode over time, leading to a gradual blurring of the memory until it eventually fades away. Right now, this aspect of memory is only partially replicated by our deep episodic memory model. Finally, we plan to investigate how forgetting influences representation learning within the episodic memory since forgetting may help to reduce the amount of redundant information balancing the dataset.

REFERENCES

[1] J. T. Wixted, "The Psychology and Neuroscience of Forgetting," *Annual Review of Psychology*, vol. 55, no. 1, pp. 235–269, Feb. 2004.

[2] S. T. Freedman and J. A. Adams, "Filtering Data Based on Human-Inspired Forgetting," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 41, no. 6, pp. 1544–1555, Dec. 2011.

[3] T. Asfour, M. Wächter, L. Kaul, S. Rader, P. Weiner, S. Ottenhaus, R. Grimm, Y. Zhou, M. Grotz, and F. Paus, "Armar-6: A high-performance humanoid for human-robot collaboration in real world scenarios," *IEEE Robotics & Automation Magazine*, vol. 26, no. 4, pp. 108–121, 2019.

[4] F. Peller-Konrad, R. Kartmann, C. R. G. Dreher, A. Meixner, F. Reister, M. Grotz, and T. Asfour, "A memory system of a robot cognitive architecture and its implementation in armarx," *Robotics and Autonomous Systems*, vol. 164, pp. 1–20, 2023.

[5] N. Vahrenkamp, M. Wächter, M. Kröhnert, K. Welke, and T. Asfour, "The robot software framework armarx," *Information Technology*, vol. 57, no. 2, pp. 99–111, 2015.

[6] J. E. Laird, "An Analysis and Comparison of ACT-R and Soar," Jan. 2022, arXiv:2201.09305.

[7] N. Derbinsky and J. E. Laird, "Effective and efficient forgetting of learned knowledge in Soar's working and procedural memories," *Cognitive Systems Research*, vol. 24, pp. 104–113, Sept. 2013.

[8] T. Williams, T. Johnson, W. Culpepper, and K. Larson, "Toward Forgetting-Sensitive Referring Expression Generationfor Integrated Robot Architectures," July 2020, arXiv:2007.08672.

[9] H.-D. Bui, T. L. Q. Dang, and N. Y. Chong, "Robot Social Emotional Development through Memory Retrieval," in *7th International Conference on Robot Intelligence Technology and Applications (RiTA)*, Nov. 2019, pp. 46–51.

[10] M. Sigalas, M. Maniadakis, and P. Trahanias, "Time-Aware Long-term Episodic Memory for Recurring HRI," in *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. Association for Computing Machinery, 2017, pp. 287–288.

[11] I. Juvina, O. Larue, and A. Hough, "Modeling valuation and core affect in a cognitive architecture: The impact of valence and arousal on memory and decision-making," *Cognitive Systems Research*, vol. 48, pp. 4–24, May 2018.

[12] K. Huang, X. Ma, R. Song, X. Rong, X. Tian, and Y. Li, "A self-organizing developmental cognitive architecture with interactive reinforcement learning," *Neurocomputing*, vol. 377, pp. 269–285, Feb. 2020.

[13] J. M. J. Murre and J. Dros, "Replication and analysis of ebbinghaus' forgetting curve," *PLOS ONE*, vol. 10, no. 7, pp. 1–23, 07 2015.

[14] C. Gurrin, H. Lee, and J. Hayes, "iForgot: A model of forgetting in robotic memories," in *5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Mar. 2010, pp. 93–94.

[15] W. Wang, B. Subagdja, A.-H. Tan, and J. A. Starzyk, "Neural modeling of episodic memory: Encoding, retrieval, and forgetting," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 10, pp. 1574–1586, 2012.

[16] J. R. Anderson, D. Bothell, M. D. Byrne, S. Douglass, C. Lebiere, and Y. Qin, "An integrated theory of the mind," *Psychological Review*, vol. 111, no. 4, pp. 1036–1060, 2004.

[17] J. E. Laird, A. Newell, and P. S. Rosenbloom, "Soar: An architecture for general intelligence," *Artificial Intelligence*, vol. 33, no. 1, pp. 1–64, 1987.

[18] J. Laird, *The Soar Cognitive Architecture*. MIT Press, 2019.

[19] S. Franklin, T. Madl, S. D'Mello, and J. Snaider, "Lida: A systems-level architecture for cognition, emotion, and learning," *IEEE Transactions on Autonomous Mental Development*, vol. 6, no. 1, pp. 19–41, 2014.

[20] G. A. Miller, "The magical number seven, plus or minus two: Some limits on our capacity for processing information," *Psychological Review*, vol. 63, no. 2, pp. 81–97, 1956.

[21] D. Vernon, *Artificial Cognitive Systems: A Primer*. The MIT Press, 2014.

[22] M. Beetz, G. Kazhoyan, and D. Vernon, "The cram cognitive architecture for robot manipulation in everyday activities," 2023.

[23] K. Kawamura, S. M. Gordon, P. Ratanawasd, E. Erdemir, and J. F. Hall, "Implementation of cognitive control for a humanoid robot," *International Journal of Humanoid Robotics*, vol. 05, no. 04, pp. 547–586, 2008.

[24] T. J. Ricker, E. Vergauwe, and N. Cowan, "Decay theory of immediate memory: From brown (1958) to today (2014)," *Quarterly Journal of Experimental Psychology*, vol. 69, no. 10, pp. 1969–1995, 2016.

[25] J. Rothfuss, F. Ferreira, E. E. Aksoy, Y. Zhou, and T. Asfour, "Deep episodic memory: Encoding, recalling, and predicting episodic experiences for robot action execution," *IEEE Robotics and Automation Letters (RA-L)*, vol. 3, no. 4, pp. 4007–4014, 2018.

[26] L. Bärmann, F. Peller-Konrad, S. Constantin, T. Asfour, and A. Waibel, "Deep episodic memory for verbalization of robot experience," *IEEE Robotics and Automation Letters (RA-L)*, vol. 6, no. 3, pp. 5808–5815, 2021.

[27] I. Tolstikhin, O. Bousquet, S. Gelly, and B. Schoelkopf, "Wasserstein auto-encoders," *arXiv preprint arXiv:1711.01558*, 2017.