Universität Hamburg
Department Informatik
Knowledge Technology, WTM

Learning to Forget in Self-Organizing Memory

Seminar Paper

Bio-Inspired Artificial Intelligence

Vadym Gryshchuk
Matr.Nr. 6435479
2gryshch@informatik.uni-hamburg.de

19.12.2018

Abstract

Lifelong learning without catastrophic forgetting is one of the main challenges of artificial agents operating in changing environments. Self-organizing neural architectures, which grow when the input distribution changes, learn spatiotemporal connections of the provided input data. During incremental tasks, where new data is provided for the agent incrementally, the networks with self-organization grow. The networks with too many nodes may not be applicable for systems, which are computationally or timely limited. Therefore, an appropriate strategy is required for removing neurons. We propose a method to remove efficiently some neurons without catastrophic forgetting considering the habituation. The conducted experiments show the efficiency of the proposed method for the neuron elimination without catastrophic forgetting.

1 Introduction

Learning without forgetting gains a rapid interest in the area of autonomous assistive robotic agents and applications. The ability to accommodate new knowledge without interference with previous learned experience is called lifelong learning. The proposed architectures for lifelong learning [10], [11], [14] are highly inspired by the processing of the information in the brain and can grow when the input distribution changes.

Growing Dual Memory (GDM) network [11] for lifelong learning of spatiotemporal representations achieves state-of-the-art results on the CORe50 dataset [7]. The GDM network utilizes two hierarchically arranged recurrent Grow When Required networks [10]. The first network, which is called growing Episodic Memory (G-EM) learns a fine-grained representation of the input data, the instances of the classes. The second network, growing Semantic Memory (G-SM), creates a compressed representation of the data learned by Episodic Memory on the category level (classes).

The recent studies in neuroscience on the Complementary Learning Systems [6], [8], [12] show that the hippocampus and the neocortex, parts of the human brain, play an important role in the recognition memory. Hippocampus is responsible for learning the structure of the received information on the very detailed level, modelling the short-term memory. Neocortex generalizes the information received from the hippocampus, where one neuron uses the overlapping representations from the hippocampus to learn high-level features.

The neurogenesis in hippocampus causes the forgetting of previous memories, but at the same time it leads to better pattern separation [3], [1]. The decay of the synapses, limitation of the synaptic strength and overwriting of the previous memories can be used to model the forgetting mechanism [13]. Forgetting is closely related to intransigence [2], the incapability of the model to learn new information, due to the restrictions that forbid to overwrite the previous knowledge.

The response of the neurons to an input can be modelled on the level of a

neuron activation. The habituation, the diminishing of a response to a repeated stimulus, is used in the GDM network to express the firing count of a neuron [11]. After multiple series of a repeated stimulus, the response of a neuron decreases [4], which leads to better learning of the input representations. A neuron has a large habituation value, when the input was provided seldom.

In this paper, we propose a method for the removal of neurons in the GDM network without catastrophic forgetting. The proposed method uses the habituation value of a neuron to efficiently remove a neuron without losing the previously learned information by the network. For this purpose, we eliminate the neuron, which responded rarely to the input after each learning phase. We evaluate the performance of the model on the CORe50 [7] dataset. We show that the integrated method for the removal of neurons efficiently reduces the size of the model.

2 Proposed Method

The proposed method for the removal of some neurons considers the habituation of the neurons. Each neuron (unit) records how often it was activated for each input sample. A low value of the habituation represents a frequent response, while the high value shows a rare activation of a neuron. The neurogenesis in the GDM network is controlled by the activation of the network and the habituation of neurons. The activation is represented through the distance between the best matching unit (BMU) and the input. If the activation of the network a is below the given threshold a_T and the habituation of the BMU h_f is less than h_T, then a new neuron will be added to the network.

During incremental learning the network decides if a new neuron should be created for each presented sample based on a and h_f. Therefore, the number of created neurons per class label depends on the variety of presented objects of the given class. If an object undergoes different visual transformations, the number of neurons grows to better represent the input. This leads to an extensive creation of neurons in the G-SM and especially G-EM. Hence, the neurons which were activated seldom will be removed and the neurons with lower habituation value will be preserved. Furthermore, the replay of the previously learned representations by Episodic Memory, to enhance memory consolidation and overcome catastrophic forgetting, causes considerable neurogenesis.

2.1 Neuron Elimination without Memory Replay

Episodic Memory is characterized by the learning of spatiotemporal representations of the input data on a fine-grained level. In Semantic Memory the neurons are formed by overlapping of neural representations from Episodic Memory. The a_T, which is usually represented by a larger value for Episodic Memory than for Semantic Memory regulates the neurogenesis. The number of neurons will grow during incremental tasks, particularly in Episodic Memory [11], leading to a slower

performance of the model in real robotic scenarios. Therefore, we define the elimination of a neuron as a threshold function of the habituation:

$$v = \mu(H) + \sigma(H), \tag{1}$$

where H is the vector representation for the habituation of all neurons in the network, μ is the mean function, σ is the standard deviation. If a habituation of a neuron is greater than v, this neuron will be removed. We define the neurons, which are removed after applying this function as weak neurons. The elimination of neurons occurs after each learning phase, where an agent has learnt the given number n of different objects or after some specific time t. In the GDM network the learning phase for CORe50 comprises all objects of one category.

The substantial difference between Episodic and Semantic memories lies in their ability to generalize input representations. Episodic Memory should encode the input on a low-level basis and be denser. Consequently, the isolated neurons that are not connected to the other neurons should not be removed. Though, the isolated neurons can be considered as outliers in the topological representation of the data, they can be activated for rare objects. Semantic Memory preserves higher grade on sparseness, but the neurogenesis is slow, and the elimination of isolated neurons can degrade the accuracy of object recognition.

2.2 Neuron Elimination with Memory Replay

Memory Replay mimics in the natural way, how the acquired knowledge is consolidated. This process occurs in human beings during sleep [9]. In the GDM network Memory Replay is modelled through the reactivated neural activity trajectories (RNATs) [11] from G-EM, which are replayed to G-EM and G-SM. This process alleviates the catastrophic forgetting by integrating the previous experiences. Each neuron in G-EM and G-SM has a number K of of temporal context descriptors, the weights of activated neurons at previous time steps. The size λ of an RNAT depends on the size of temporal context descriptors and is defined as $\lambda = K^{EM} + K^{SM} + 1$

After each learning phase the neural trajectories of all neurons in the network are replayed. It is important that RNATs update the weights of neurons, which are the potential candidates for the interference with new knowledge. The creation of new neurons must be limited due to the consolidation of previous knowledge by Memory Replay. Consequently, we define the threshold for the elimination of neurons in the GDM network with Memory Replay as follows:

$$z = e^{-\mu(H)}, \tag{2}$$

where e is the base of the natural logarithm. The threshold $z \in [0, 1]$ conveys the importance of preserving the neurons. If $\mu(H)$ is low, which means that neurons respond frequently to stimulus, then z will be large and only a small number of neurons will be deleted for better knowledge preservation. Otherwise, if $\mu(H)$ is

large, which leads to rare responses of neurons, then z will be low and more neurons will be deleted.

3 Experiments and Results

We conduct experiments to eliminate some neurons with the proposed method. We use the CORe50 dataset for the analysis and evaluation. The incremental learning strategy is used to test our approach. The code is available at a repository[1].

3.1 Dataset

The CORe50 dataset [7] consists of 50 objects divided into 10 categories. Each object has ca. 300 image sequences recorded with different visual transformations (e.g. lighting conditions, background). There are 5 objects in each category. We have reduced the number of frames per second to 5, which is the same as in [11]. There are 11 sessions that were used to collect the data. Three sessions were recorded outside, the other 8 sessions were taken indoor. For the training we use only one session with the number 1 and for the testing the session with the number 3. The main aspect is not to compare the results with [11], [7], but to judge the results of the model with and without the proposed method for neuron elimination. For this purpose we consider experiments on the GDM with and without Memory Replay, where we compare the results of our proposed method to the results that can be achieved by the original implementation of the GDM[2]. We have used the ResNet-50 model [5] as a feature extractor.

During incremental learning the data becomes progressively available for training. We define 10 learning phases. Each learning phase comprises all data of one category. We use the same training parameters listed in [11].

3.2 Incremental Learning without Memory Replay

We evaluate at first the incremental learning strategy without Memory Replay. Fig. 1 reports the number of neurons of the G-SM network for each category. In Fig. 1 (left) the neurons are not removed. We can see that the model has created a lot of neurons for the second category. In Fig. 1 (right) neuron elimination method has reduced the number of neurons for this category extensively. The association frequency of each neuron in this category with the given label may differ. Fig. 2 shows the number of activations of each neuron in the same category. The removal of weak neurons reduced the number of neurons notably. Only 4 neurons were created, meanwhile 29 neurons were generated without the neuron elimination method.

[1]The GDM network with the proposed method for the elimination of neurons: `https://github.com/VadymV/GDM`

[2]Original implementation of the GDM network: `https://github.com/giparisi/GDM` accessed 27 January 2019.

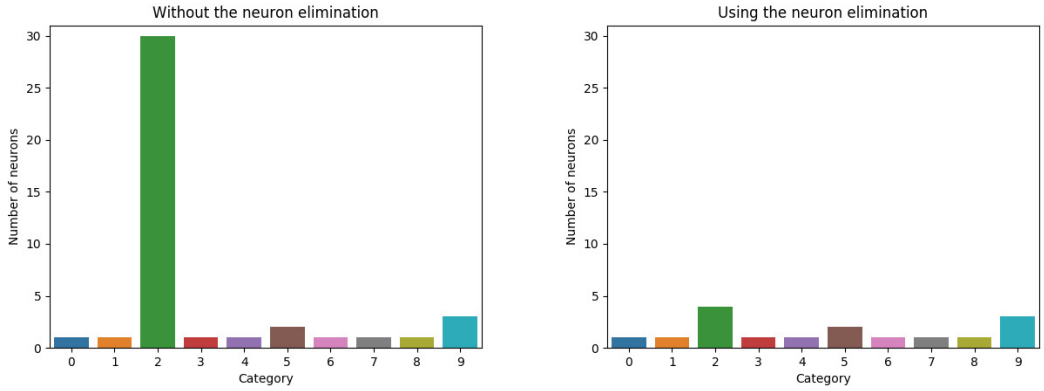

Figure 1: The number of neurons of the G-SM network for each category without neuron elimination (left) and with neuron elimination (right).

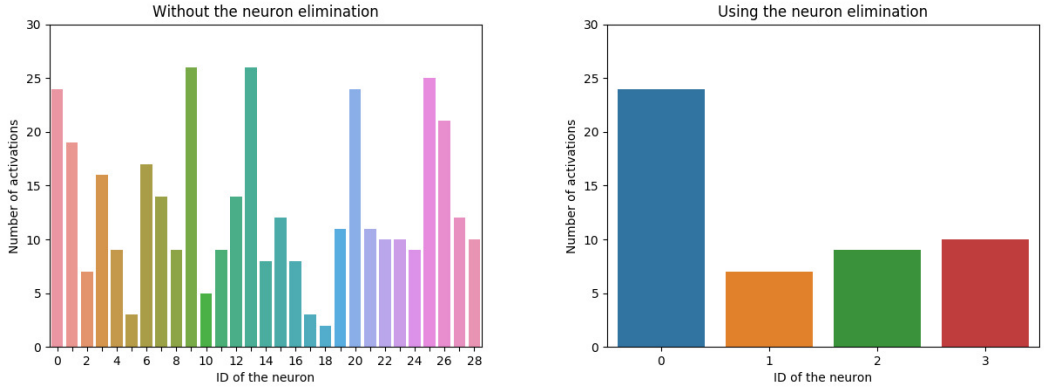

Figure 2: The number of activations for each neuron of the G-SM network for the samples of one category without (left) and with (right) neuron elimination.

Fig. 3 reports the number of created neurons during incremental learning of each category. The number of neurons has increased in the original implementation of the GDM network, while the elimination method removed most of the neurons associated with the second category in G-SM, as shown in Fig. 2. In G-EM the number of neurons grows in both implementations, but the removal method has minimized the total number of neurons.

Fig. 4 shows the overall classification accuracy for G-SM and G-EM without neuron elimination and with neuron elimination. The removal of weak neurons has not decreased the accuracy of both networks. Both implementations show the same behaviour of classification accuracies.

3.3 Incremental Learning with Memory Replay

In this subsection we evaluate the incremental learning strategy with Memory Replay, where RNATs are replayed to G-EM and G-SM. The length of an RNAT

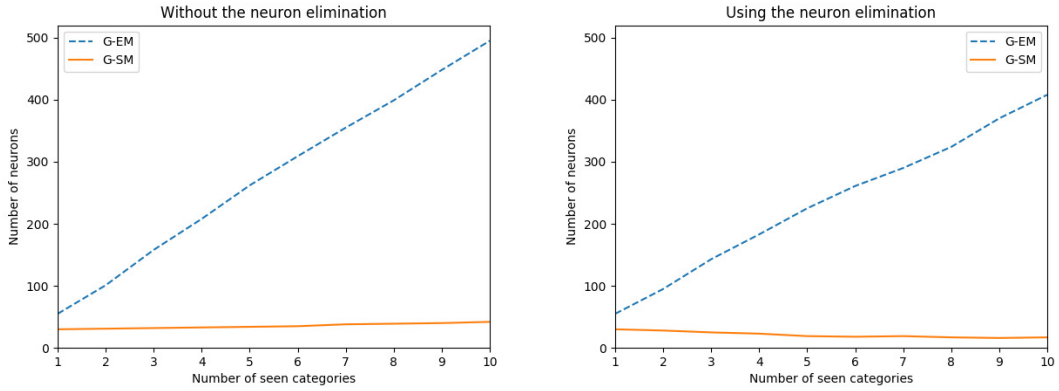

Figure 3: The number of neurons during incremental learning without (left) and with (right) neuron elimination.

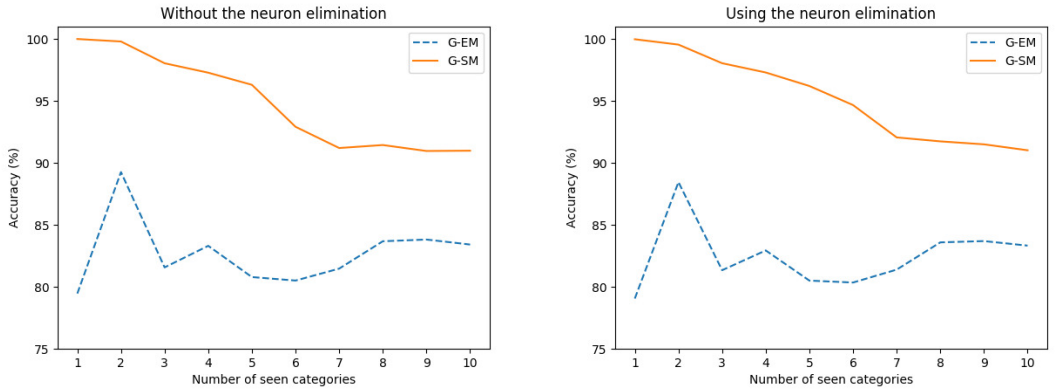

Figure 4: The overall classification accuracy of both networks using incremental learning strategy without (left) and with (right) neuron elimination.

is computed as in [11]. When the number of temporal context descriptors is set to 2 for both networks, the length of the RNAT is 5. The RNAT of each neuron in the G-EM network consists of the sequence of the consecutively activated neurons, which have the greatest temporal synaptic link [11].

Fig. 5 reports the number of created neurons for the samples of each incrementally available category. It is noticeable, that the neuron elimination method removed most of the neurons during the replay step, particularly in G-EM. The implementation without neuron removal adds a high number of new neurons in the G-EM network during each replay phase. This is caused by the replay of the all RNATs after each learning phase. Therefore, the replay sequences get larger with each step. The model judges many of the replay sequences as novel objects and creates new neurons. The maximum allowed number of neurons in the model is set to 10K, which is not suitable for incremental learning, but for this task it should be sufficient. In Fig. 5 (left) it can be seen, that no further neurons are allowed to be created in G-EM, when the model learns the ninth category.

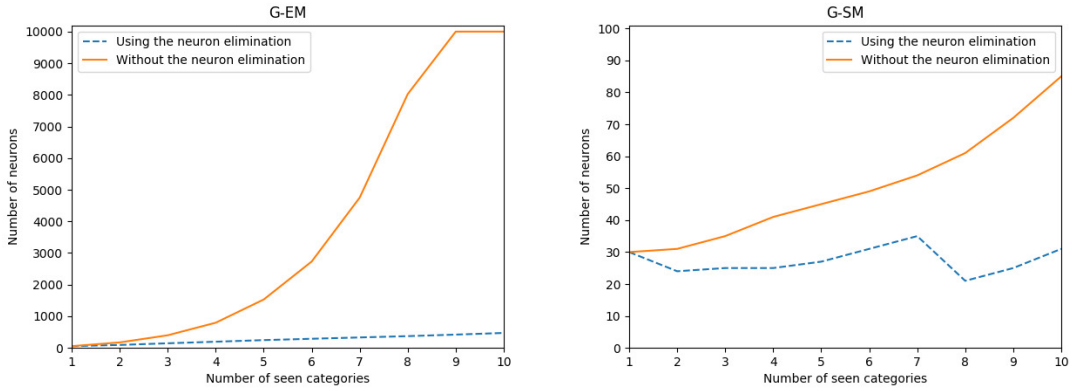

Figure 5: The number of neurons during incremental learning without neuron elimination (left) and with neuron elimination (right) of the trained model with the Memory Replay.

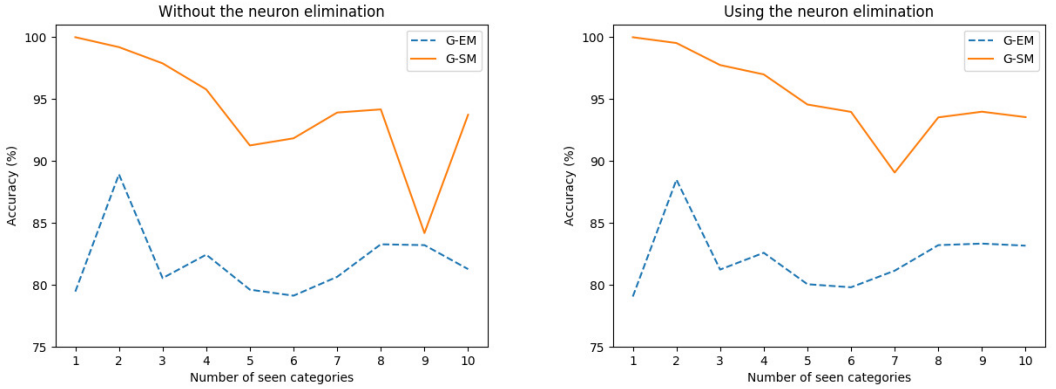

Figure 6: The overall classification accuracy of G-EM and G-SM without neuron elimination (left) and with neuron elimination (right) of the trained model with Memory Replay.

Fig. 6 reports the overall accuracy. The results of both implementations are not significantly different and show the same trend. In Fig. 6 (left) it can be seen, that the accuracy of G-EM drops for the last category. It is explained by the maximum number of allowed neurons the model has reached during the previous learning phase and no new neuros were created afterwards. The model with Memory Replay gained better results for G-SM in contrary to the model without replay.

Fig. 7 shows the classification accuracies of the first category. The trained models with and without Memory Replay achieve approximately the same results, apart from the G-SM network, when it encounters the last category. The difference of results can be explained by the great resolution of replay data shown to the G-SM network that was trained without neuron elimination. The maximum allowed number of neurons causes the same effect as in Fig. 6 (left) and the accuracy of the G-EM network drops at the end.

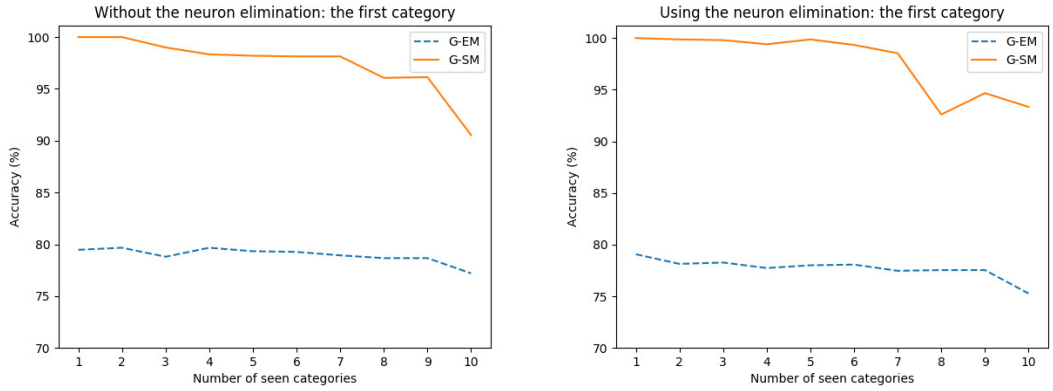

Figure 7: The classification accuracy of the objects of the first encountered category during incremental learning without neuron elimination (left) and with neuron elimination (right) of the trained model with Memory Replay.

4 Discussion

We proposed a method to remove neurons without catastrophic forgetting. We defined a threshold for the elimination of a neuron in G-EM and G-SM networks. This threshold is based on the habituation of each neuron and expresses the diminishing response to a frequent stimulus. The comparison results show a little negligence of the classification accuracy of the first seen objects. Nevertheless, the reduction of the size of the model delivers promising results. Though, the evaluation is limited in the sense of incremental tasks the agent learns. Real scenario would include many iterative learning phases, which could span for a larger time.

Though, the proposed method achieves good results, it is not evaluated on the scenarios when the number of samples per class is not uniformly distributed, which could be the possible future work. Unbalanced training data may lead to the omitting of the input information, when, for example, an object is encountered only once during the learning phase. This behaviour is caused by the principle of removing those neurons, that are very seldom activated. Though, this mechanism models the forgetting in the human brain, an artificial agent would benefit from remembering everything, but a consensus due to computation time and hardware limits must be met.

5 Conclusion

In this paper we proposed a method for the elimination of neurons considering their habituation. The conducted experiments showed that the proposed method can achieve good results in comparison to the original implementation and does not cause the problem of catastrophic forgetting. We identified the challenges and possible future work to enable the model learn better representations of the input data.

6 Appendix

The code is available at a repository: `https://github.com/VadymV/GDM`.
The extracted features from the CORe50 dataset can be found here: `https://drive.google.com/open?id=1vO9SLdhg4iMsRIB1ntHzVeVrKgxA32wr`

References

[1] Katherine G. Akers, Alonso Martinez-Canabal, Leonardo Restivo, Adelaide P. Yiu, Antonietta De Cristofaro, Hwa-Lin (Liz) Hsiang, Anne L. Wheeler, Axel Guskjolen, Yosuke Niibori, Hirotaka Shoji, Koji Ohira, Blake A. Richards, Tsuyoshi Miyakawa, Sheena A. Josselyn, and Paul W. Frankland. Hippocampal neurogenesis regulates forgetting during adulthood and infancy. *Science*, 344(6184):598–602, 2014.

[2] Arslan Chaudhry, Puneet Kumar Dokania, Thalaiyasingam Ajanthan, and Philip H. S. Torr. Riemannian walk for incremental learning: Understanding forgetting and intransigence. In *Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part XI*, pages 556–572, 2018.

[3] Paul W. Frankland, Stefan Khler, and Sheena A. Josselyn. Hippocampal neurogenesis and forgetting. *Trends in Neurosciences*, 36(9):497 – 503, 2013.

[4] Catharine H. Rankin, Thomas Abrams, Robert Barry, Seema Bhatnagar, David Clayton, John Colombo, Gianluca Coppola, Mark Geyer, David Glanzman, Stephen Marsland, Frances Mcsweeney, Donald Wilson, Chun-Fang Wu, and Richard Thompson. Habituation revisited: An updated and revised description of the behavioral characteristics of habituation. *Neurobiology of learning and memory*, 92:135–8, 10 2008.

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 770–778, 2016.

[6] Dharshan Kumaran, Demis Hassabis, and James L. McClelland. What learning systems do intelligent agents need? Complementary learning systems theory updated. *Trends in Cognitive Sciences*, 20(7):512 – 534, 2016.

[7] Vincenzo Lomonaco and Davide Maltoni. CORe50: a new dataset and benchmark for continuous object recognition. In *1st Annual Conference on Robot Learning, CoRL 2017, Mountain View, California, USA, November 13-15, 2017, Proceedings*, pages 17–26, 2017.

[8] Kenneth A. Norman. How hippocampus and cortex contribute to recognition memory: revisiting the complementary learning systems model. *Hippocampus*, 20 11:1217–27, 2010.

[9] Freyja H. Olafsdottir, Daniel Bush, and Caswell Barry. The role of hippocampal replay in memory and planning. *Current Biology*, 28(1):R37 – R50, 2018.

[10] German I. Parisi, Jun Tani, Cornelius Weber, and Stefan Wermter. Lifelong learning of human actions with deep neural network self-organization. *Neural Networks*, 96:137–149, 2017.

[11] German I. Parisi, Jun Tani, Cornelius Weber, and Stefan Wermter. Lifelong learning of spatiotemporal representations with dual-memory recurrent self-organization. *CoRR*, abs/1805.10966, 2018.

[12] Alison R. Preston and Howard Eichenbaum. Interplay of hippocampus and prefrontal cortex in memory. *Current Biology*, 23:R764–R773, 2013.

[13] Edmund Rolls. *Cerebral Cortex*. Oxford University Press, 2016.

[14] Boaz Vigdor and Boaz Lerner. The Bayesian ARTMAP. *IEEE Transactions on Neural Networks*, 18(6):1628–1644, Nov 2007.