Memory modification from perspective of Free Energy Principle (FEP)

Dong-Hui Song

State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University

Email: donghuipsy@gmail.com

Abstract

Memory can be modified, free energy principle as a unified theory of brain can explain memory modification. Agent to modify original memory or create new memory to minimize free energy. Memory modification changes the synaptic plasticity rather than the facts that are stored in intracellular molecular. Memory modification also is a procedure of learning, maybe real intelligence must have a body to get sensory and act environment to reduce free energy which can change memory and create new memory.

Introduction

Memory is an important function of brain in human and animal, which can store what they learn for future use to implement more complexity function for adaptive complicated environment. However, memory is never immutable, on the contrary, memory can be modified immediately after retrieval (Gershman, Monfils et al. 2017). What caused the memory modification? Gershman at al (2017) developed a computational theory of memory modification, central to the theory is memory is inferential in nature: decisions about when to modify an old memory or form a new memory are guided by inferences about the latent causes of sensory data (Gershman, Radulescu et al. 2014, Gershman, Norman et al. 2015). More specifically, according to which modification of a memory trace occurs through classical associative learning, but which memory trace is eligible for modification depends on a structure learning mechanism that discovers the units of association by segmenting the stream of experience into statistically distinct clusters (latent causes). New memories are formed when the structure learning mechanism infers that a new latent cause underlies current sensory observations. By the same token, old memories are modified when old and new sensory observations are inferred to have been generated by the same latent cause (Gershman, Monfils et al. 2017). The computational framework was derived from probabilistic principles, which consist with Bayesian brain hypothesis (Knill and Pouget 2004). However, Bayesian brain hypothesis is subsumed by the free energy principle (FEP) under a larger view (Friston, Kilner et al. 2006). In this review, I will try to explain memory modification from perspective of FEP which is regarded as a unified theory of brain (Friston 2010).

Free Energy Principle

Free energy arises in many contexts, especially physics and chemistry. In thermodynamics, free energy is a measure of the amount of work that can be extracted from a system. In 2006, Karl·Friston proposed a FEP for the brain, which starting with the notions of neuronal energy described by Helmholtz (1860) and more recent formulations like Bayesian inversion and predictive coding (Dayan, Hinton et al. 1995, Rao and Ballard 1999). Karl·Friston indicated this free energy is not a thermodynamic free energy, but a free energy formulated in terms of information theoretic quantities. Free energy is a function of a recognition density and sensory input. The free energy considered here measures the difference between the probability distribution of environmental quantities that act on the system and an arbitrary distribution encoded by its configuration. The system can minimize free energy by changing its configuration to affect the way it samples the environment or change the distribution it encodes. These changes correspond to action and perception respectively and lead to an adaptive exchange with the environment that is characteristic of biological systems. This treatment assumes that the system's state and structure encode an implicit and probabilistic model of the environment. The FEP says that any self-organizing system that is at equilibrium with its environment must minimize its free energy (Friston 2010). In a nutshell, FEP is that the brain seeks to minimize surprise (Friston, Kilner et al. 2006, Gershman 2019), to minimize surprise or free energy from two ways, one is to change the extent environment (sampling the environment or action), another is to change the internal state (the distribution it encodes or perception).

Computational model of memory modification and FEP

Computational model of memory modification from Gershman (2017), which regards memory modification from combine of structure learning that infer latent causes using Bayes' rule and associative learning that update weights using the delta rule (Fig.1). When prediction errors are small, the posterior probability of the acquisition latent cause is high then leading to modification of the original memory, if prediction errors are very large, the posterior probability of the acquisition latent cause is low then leading to formation of a new memory. According to FEP, memory modification or create new memory is a procedure of the brain seeks to minimize surprise or minimize the predictive error. When agent in a new environment or event, new current sensory observations are inputted, the brain according to existed memory (prior probability) to create the posterior probability to predictive the new environment or event, if the predictive error is small, original memory is modified, but if the predictive error is large, agent is going to resampling the environment or act the environment then create the new memory to reduce predictive error, in another word, minimize the surprise.

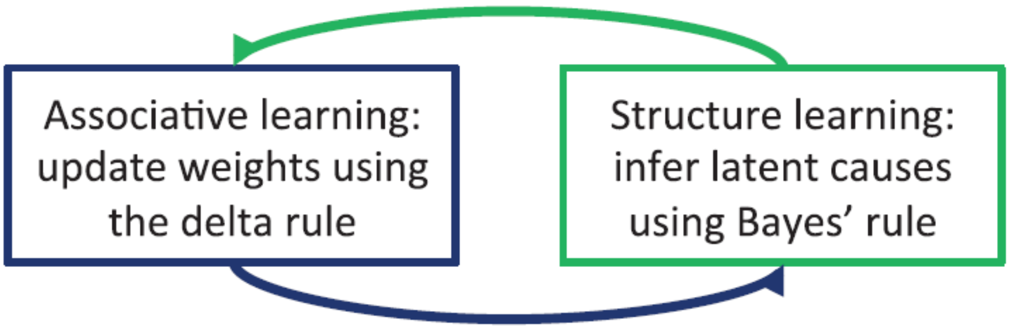

Fig.1 A high-level schematic of the computations in the latent-cause model. Associative learning, in which the associative weights are updated (using the delta rule) conditional on the latent-cause posterior, alternates with structure learning, in which the posterior is updated (using Bayes' rule) conditional on the weights (Gershman, 2017).

What is changed when memory is modified?

Understanding the memory modification, we must know what are content and structure of memory. Content refers to the information encoded in memory. Structure refers to the code that maps information into a physical trace (or “engram”) (Gershman 2021). The prevailing view is that the content of memory consists of associations between events; the code is a mapping from associations to synaptic strengths from standard textbook (Martin, Grimwood et al. 2000, Takeuchi, Duszkiewicz et al. 2014). However, memories are encoded in an intracellular molecular substrate was recently revived (Abraham, Jones et al. 2019) that originally proposed in the mid-20th century (Gaito 1976). The content of memory consists of “facts” rather than associations under this view(Abraham, Jones et al. 2019). These facts are read out by the spiking activity of neurons (presumably mediated by some biochemical process within the cell) and thereby made computationally accessible to downstream neurons. Synaptic plasticity, on this view, plays no role in memory storage(Abraham, Jones et al. 2019). Currently, Gershman (2021) developed a model that an intracellular molecular mechanism stores “facts”—in particular, the parameters of a generative model. To do some useful computation with these facts (e.g., perception, prediction, action), the brain needs to infer the latent causes of sensory observations by inverting the generative model. This is generally intractable for complex generative models but can be approximated using a parametrized mapping from observations to a distribution over latent causes—the inference model. He hypothesizes that: (1) this inference model is implemented by spiking activity in a network of neurons;(2) its parameters are stored at the synapse; (3) synaptic plasticity updates these parameters to optimize the inference model. According to FEP and Gershman’s model, memory modification is changes of the synaptic plasticity (association) rather than the facts that are stored in intracellular molecular. Memory modification or create new memory is minimize the surprise or free energy procedure. Our brain is very efficient, if it is efficient, then it is impossible to spend a lot of energy to encode a large amount of fact information in daily life, and an efficiently the encoding of information should be a difference coded to reconfigure (synaptic plasticity) those basic facts already exist that are stored in intracellular molecular. This explains why episodic memory is more efficient, while semantic memory is relatively inefficient, and episodic memory is easier to change while semantic memory is less prone to change. Because we already store the facts and patterns of synaptic connections in our episodic memory, reconfigure of the facts and patterns is a relatively fast process, but in semantic memory, although we have stored some basic facts, such as the meaning of each word, its shape, color, etc., but we need to generate more synaptic activity to make new connections for semantic.

Conclusions and future directions

To reduce surprise, we have to change sensory input, it says that our actions should also minimize free energy. We are open systems in exchange with the environment; the environment acts on us to produce sensory impressions and we act on the environment to change its states. This exchange rests upon sensory and effector organs (like photoreceptors and oculomotor muscles). If we change the environment or our relationship to it, sensory input changes. Therefore, action can reduce free-energy by changing sensory input, whereas perception reduces free-energy by changing predictions (Friston, Kilner et al. 2006, Friston 2010). Memory modification is essentially a reduce free energy (prediction errors) procedure, agent can constantly to modify memory or create new memory by changing sensory input and acting on the environment to reduce the free energy. Memory modification implement by changing the synaptic plasticity (association) rather than the facts that are stored in intracellular molecular.

Memory modification also is a procedure of learning, maybe real intelligence must have a body to get sensory input and action to reduce free energy, which can change memory and create new memory. Now artificial intelligence and machine learning are still very hot in academic and industry from where Hinton gained a huge advantage in ImageNet classification with deep convolutional neural networks (Krizhevsky, Sutskever et al. 2012), but we are still far from real intelligence. We more pay attention to the brain but is not an integrated agent which with sensory input, computation, reference, action to reduce the free energy. By continuously feeding the neural network with big data, it is always limited or even impossible, especially, it may lose its effect when it encounters a new environment. Real intelligence need to active inference (Friston, Lin et al. 2017) , embodied AI maybe is available to realize the real intelligence, current works shown advances of embodied AI (Gupta, Savarese et al. 2021, Nygaard, Martin et al. 2021).

Finally, FEP involves a lot of theory from disciplines, especially physic, it seems to indicate if human and the universe have a same source, human as a part of the universe, therefore physic rules are suit to human and neuroscience.

Reference

Abraham, W. C., O. D. Jones and D. L. Glanzman (2019). "Is plasticity of synapses the mechanism of long-term memory storage?" NPJ science of learning **4**(1): 1-10.

Dayan, P., G. E. Hinton, R. M. Neal and R. S. Zemel (1995). "The helmholtz machine." Neural computation **7**(5): 889-904.

Friston, K. (2010). "The free-energy principle: a unified brain theory?" Nature reviews neuroscience **11**(2): 127-138.

Friston, K., J. Kilner and L. Harrison (2006). "A free energy principle for the brain." Journal of physiology-Paris **100**(1-3): 70-87.

Friston, K. J., M. Lin, C. D. Frith, G. Pezzulo, J. A. Hobson and S. Ondobaka (2017). "Active inference, curiosity and insight." Neural computation **29**(10): 2633-2683.

Gaito, J. (1976). "Molecular psychobiology of memory: Its appearance, contributions, and decline." Physiological Psychology **4**(4): 476-484.

Gershman, S. J. (2019). "What does the free energy principle tell us about the brain?" arXiv preprint arXiv:1901.07945.

Gershman, S. J. (2021). "The molecular memory code and synaptic plasticity: a synthesis."

Gershman, S. J., M.-H. Monfils, K. A. Norman and Y. Niv (2017). "The computational nature of memory modification." Elife **6**: e23763.

Gershman, S. J., K. A. Norman and Y. Niv (2015). "Discovering latent causes in reinforcement learning." Current Opinion in Behavioral Sciences **5**: 43-50.

Gershman, S. J., A. Radulescu, K. A. Norman and Y. Niv (2014). "Statistical computations underlying the dynamics of memory updating." PLoS computational biology **10**(11): e1003939.

Gupta, A., S. Savarese, S. Ganguli and L. Fei-Fei (2021). "Embodied Intelligence via Learning and Evolution." arXiv preprint arXiv:2102.02202.

Knill, D. C. and A. Pouget (2004). "The Bayesian brain: the role of uncertainty in neural coding and computation." TRENDS in Neurosciences **27**(12): 712-719.

Krizhevsky, A., I. Sutskever and G. E. Hinton (2012). "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems **25**: 1097-1105.

Martin, S. J., P. D. Grimwood and R. G. Morris (2000). "Synaptic plasticity and memory: an evaluation of the hypothesis." Annual review of neuroscience **23**(1): 649-711.

Nygaard, T. F., C. P. Martin, J. Torresen, K. Glette and D. Howard (2021). "Real-world embodied AI through a morphologically adaptive quadruped robot." Nature Machine Intelligence **3**(5): 410-419.

Rao, R. P. and D. H. Ballard (1999). "Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects." Nature neuroscience **2**(1): 79-87.

Takeuchi, T., A. J. Duszkiewicz and R. G. Morris (2014). "The synaptic plasticity and memory hypothesis: encoding, storage and persistence." Philosophical Transactions of the Royal Society B: Biological Sciences **369**(1633): 20130288.