How neural networks give us an insight into dreaming.

Avinash Dindial – 1006020949

Abstract

Dreams probably have no function, but they do have coherence and meaning, which is often conflated with function; it is a by-product of the evolution of sleep and consciousness (Domhoff, 2003). Although the science of sleep is something relatively well understood, the feature of dreaming during sleep is still shrouded in mystery. Neural networks attempt to mimic the human brain in its processes and pattern recognition skills. It is known that deep neural networks (DNN) face the problem of overfitting as they learn. Two ways to combat this to introduce noise or using the drop out technique. These techniques afford the network unsupervised plasticity which is the very thing that is going on in the brain. This paper seeks to show that the brain faces a similar problem of overfitting, as we learn new things every day and there are constantly large amounts of inputs sent to our brain at any given time. I argue that dreams are the way in which the brain introduces noise to make the brain self-organize itself to make learning more efficient, drawing inspiration from Erik Hoel and his Overfitted Brain Hypothesis. Crick and Mitchinson (1983, 1995) propose that dreams may be a natural result of neural net overloading in the CNS. Sleepless nights, in particular nights void of REM sleep, can lead to problems in memory among other cognitive issues. There seems to be a link between cognitive decline and lack of REM sleep, which will be further discussed in this paper. Although different in execution, the way the brain and neural networks deal with overfitting is strikingly similar. Due to these similarities, neural networks may provide us with insights about

the brain that we have never had before. It is the hope of this paper that some light can be shed on the phenomenon of dreaming and to highlight its importance with the help of neural networks.

Introduction

The fact that all animals sleep suggests that it is a vital part of life, needed for survival. While sleep can be explained as a body's need to rest, or rebuild muscle tissue etc., the feature of dreaming is still up for debate. Some argue that dreams are useless, while others argue it is essential for mental and physical health. We cannot actively test dreams as we cannot go into a person's mind to see what is going on but with the advent of neural networks and computers, we can simulate a brain to get some answers. Dreams occur during the REM cycle and can become vivid, complex and may revolve around an intricate storyline. There are theories that suggest dreams may be nothing more than an evolutionary trait, designed to help early humans deal with threats by rehearsing different situations during sleep. While this is a worthwhile argument, there must be other reasons for why we dream as this evolutionary explanation does not explain why we see deficits in memory after limiting REM sleep in individuals, so there must be hidden cognitive functions at work. Neural networks, such as Google's DeepDream, can be used to visualize the phenomena of dreaming although they are very limited in what they can dream about. In the case of DeepDream, the program attempts to recognize structures in images and display them back to the user. So, if given a data set to train on of thousands of pictures of cars, it will eventually become very good at recognizing cars if shown a picture, Now, if we want the program to return a picture without giving it any input, we then allow it to “dream” up a picture for us, often it will be very strange, with lots of mistakes, but nevertheless have some likelihood to a car. This is very similar to how humans dream. If you can recall, I am sure your dreams at some point consisted of strange events, funny looking objects, people etc. So, we see a similarity here already in how neural networks and humans dream that results in warped images or other strange events. Research suggests that dreams play a role in processing the information, both consciously and unconsciously, that have been taken in during the day. This may be why some of our dream characters may seem like complete strangers to us, but they may be someone we passed in the street but paid little attention to. Similarly, during sleep, our brain comes up with images of the world, but it does not have the input from the environment; it has to create it for itself. This is why dreams can be so strange, because the brain, like the neural network, has to recreate images for itself. In addition to this, emotional areas of the brain are more active during REM sleep compared to the parts responsible for logical thinking (Stickgold, 2014). This results in emotion fuelled dreams with a narrative that would be seen as illogical in the real world, but in a dream it makes perfect sense. Suppose we have a neural network trained on recognizing pictures of women and fishes. Neural networks operate on wake and sleep phases. During the waking phase, information is coming into the network. During the sleep phase, i.e., the phase where no more information is coming into the network, the network consolidates patterns and optimizes information storage and we see a phenomenon known as artificial dreaming. This is where the network now generates variations of the invariant patterns it detected during its waking phase. These artificial dreams generate fantasies of the recognized patterns in hopes of creating a recreation of previously observed data i.e., it is self-organizing. These fantasies are not random, rather they are often hybrids or combinations of observed data. In our neural network that observed women and fishes, during an artificial dream, the formation of creatures such as mermaids, which obviously does not exist in the real world still show up in our neural networks. This all points to the fact that artificial systems like neural networks can function similar to brain processes. Therefore, neural networks can give us insight into dreaming, helping us understand for the first time, its functions and importance.

Now that we have established the similarities between artificial and natural dreams, we can now go on to breaking down some key areas in which dreams play a vital role. This brings us to the Overfitted Brain Hypothesis (OBH), proposed by Erik Hoel in 2020. This theory suggests that dreams, and all their properties are the biological equivalent of injecting noise into a neural network to prevent overfitting of data. This theory suggests that the human brain is overfitted from all the sensory input it absorbs during the day and in order to minimize overfitting by data compression, injecting noise etc., therefore allowing generalization and efficiency. Of course, overfitting can be seen in other aspects of our lives, in our everyday life, humans constantly “overfit” some subroutines via “muscle memories” (Zhang et. al., 2018). The need for generalization in the human brain is clear, we cannot process or keep all the information that we encounter on a daily basis, so we need a mechanism that filters the important information in such a way that it lends itself to be applicable in other contexts (recall learning and reconstructive memory etc.). In neural networks, overfitting refers to the issue where a network trains and fits one data set but loses the ability to generalize with others. To combat this, we use disrupting strategies such as shutting down the network or injecting noise which allows the network to self-organize. This is very similar to how incubation, or time spent away from solving a problem, can improve insight. This self-organizing can be seen both in our brains and well as computationally in neural networks. The most compelling form of evidence for this theory is intentionally triggering specific dreams about a specific topic. Through repetitive tasks, such as playing tetris all day, can lead to dreams that involve playing Tetris (Stickgold, 2010). The result of doing a repetitive task is that it overfits the brain, and during sleep, the brain attempts to generalize the game and it does this by dreaming about it. You would not dream about specific moves in particular but there is evidence that after sleep, performance on this and other cognitive tasks increase. Dream content is related to daytime experience (Freud, 1900). Similarly, the brains of those with various disorders can become overfitted with emotion or recurring/intrusive thoughts leading to nightmares, a common occurrence in most individuals with mental illness. Anxious individuals are more likely to have confused and anxiety-ridden dream content (Pagel et. al., 2017). Dreams also reflect one's interests and personality, just like mental activity during wakefulness (Yuval et. al., 2010). These claims all converge on the OBH as dreams seem to break down what is already going on in the mind and reorganize it in a way beneficial to the brain. This property of self-organizing affords plasticity to the brain. We can also see plasticity in neural networks from the change in their architecture. While both nonrapid eye movement (NREM) sleep and rapid eye movement (REM) sleep may play a role in synaptic plasticity, REM sleep in particular may be important for developmentally regulated plasticity (Meerlo et. al., 2015). Infants spend more time in REM sleep compared to older children or adults (Cao et. al., 2020). During this time of development (up until 3 years of age), sleep is used for neural development and then after that sleep is used for repair and clearance (Cao et. al., 2020). In neural networks, we observe the equivalent of plasticity by using techniques such as drop out and then we again observe the “clearance” Cao et. al. discussed as dealing with overfitting. This makes sense as babies’ brains are more plastic compared to adult’s brains. To fully appreciate how much neural networks can tell us about the brain, consider

Piaget’s distinction between qualitative and quantitative development. We can consider the waking phase of a neural network as developing quantitatively while the sleeping phase as developing qualitatively as its competence increases as a result of the change in architecture from reorganizing. The strangeness of dreams that was discussed earlier in this paper, according to Hoel, is what gives dreams their biological functions. The OBH conceptualizes dreams as a form of purposefully corrupted input, likely derived from noise injected into the hierarchical structure of the brain, causing feedback to generate warped or "corrupted" sensory input, the overall evolved purpose of this stochastic activity is to prevent overfitting (Hoel, 2020). There are a couple of techniques used to combat this issue of overfitting, each with some relation to dreaming in humans. Firstly, dropout technique, which promotes unsupervised plasticity, occurs by shutting off input nodes and therefore altering the architecture of the neural network. This is the same process that humans use. Dreams seem to resemble this technique as there is very few details (texture etc.) in dreams, there is less information compared to when awake. In this case, the less relevant information is *dropped*, resulting in the sparseness of dreams. This technique weakens the casual inferences made during the waking phase. It allows the network to explore more options and therefore increases its chance to make the correct inference. According to Hoel, this increases the salience of relevant information, such as a lesson learnt, and attempts to minimize irrelevant information, such as a person’s shoe colour. This self-organizing feature is what makes neural networks especially promising to provide us with an explanation of how insight is grounded in the brain, but that is the topic for another paper.

Secondly, the stacking of networks on top one another can replicate feedback within itself and allow for deep learning. Each network in this stack has its own wake/sleep cycle that bootstrap on each other as you go up the stack in the hopes of being able to accurately predict patterns in the world. The better a network gets at predicting itself, the better at predicting that part of itself that is interacting with the world hence it gets better than predicting the world (Vervaeke, 2021). Humans may be using dreams to better their predictive power as dreams afford us the variation that helps us predict the world. In dreaming, the processes of prediction error minimization and hypothesis testing are simulated largely internally, but nonetheless in a fairly realistic manner (Windt et. al., 2019). This avenue suggests that these layered neural networks and dreams have a top-down structural hierarchy. Top-down systems are largely self-contained and not dependent on information from other systems or sources (Foulkes & Dumhoff, 2014). Dreams are top-down (Foulkes et. al, 2014) which is consistent with brain stem activity during REM sleep. In contrast, during waking, we are constantly bombarded with sensory inputs which are of a bottom-up hierarchy. Higher networks in the hierarchy predict themselves (self-contained) and so when noise is injected, as the OBH claims, these self-predicting networks generate warped images and as already explained prevent overfitting. So we see neural networks, backed by the theory of the OBH, can provide us with two possible functions of dreaming: to increase predictive power and ultimately prevent overfitting. There is much more to this, but for the sake of this paper, I will stop at these two.

One of the most popular theories of dreams is that it is for memory consolidation. The view of this theory is that previously stored memories are strengthened by being accessed during dreams or that dreams are the result of the hybridization of new and old memories. It is clear that sleep mediates learning and memory processing, but the way in which it does so remains largely unknown (Stickgold, 2014). It has been suggested that dream content may represent the brain's effort to elaborate on recent memories via novel associations with remote memories which are emotionally salient (Llewellyn, 2013). Consistent with the theory, several studies have shown that emotional memories are particularly strengthened across sleep in particular when containing high amounts of REM sleep (Rasch, 2013). Recall that the brain and neural networks both undergo data compression to make their data more generalizable, better at being predictive and ultimately combats overfitting. Suppose subjects are given a number of different patterns and then a mathematical average of all the patterns is calculated and shown to them. Most, if not all, would reply that they have seen the pattern generated from the average, but they have never seen it before. Memory therefore undergoes data compression, allowing itself to be generalized in the same way we have discussed before. Recall the Tetris example mentioned previously. We observe persons do better on this and other tasks after sleep, so it is probably due to this data compression as the brain restructures itself to better recognize patterns. So, maybe sleep does not specifically pull memories and intertwine them with old ones, rather it generalizes; using old and new memories together to gain predictive power. In a recent paper by Fachechi et. al., they proposed a neural network capable of reinforcing important memories and getting rid of irrelevant ones. In the presence of noise, pure memories remain stable, and the network performance increases with dreaming time (Fachechi et. al.). However, this notion of forgetting certain memories is largely ignored by the scientific community, whereas the theory of consolidation is widely accepted. The consolidation theory merely states there is a correlation between sleep and memory processing but does not give us an explanation as to why this occurs. As we know, correlation is not causation, so this theory is weak. In addition, the view that memories are replayed during dreams is problematic. Reconstructive memory is known to change over time, so if you keep calling on memories during dreams, then over time these memories will change not be reinforced. This theory also states that memories are the main thing that is being replayed, manipulated etc., but we have already established that dreams warp our sense of the world. How then can these doctored memories reinforce old ones or how does having a warped sense of the day's events possibly consolidate memory? Except in the cases of PTSD, dreams are often fantastical, weird, and barely resemble anything from real life. It is therefore highly unlikely that dreams are replays of memory (Hoel, 2020). In contrast to replays of memory, a better theory would be data compression or as Erik Hoel claims, it would better fit the OBH.

In this paper, I have highlighted the OBH as a very plausible theory of dreaming, using neural networks as the backbone of my explanations. Neural networks do give us a different perspective of natural dreams by analysing artificial dreams as they behave similar. I also showed an opposing and widely accepted theory- memory consolidation. Again, using neural network theory, I have showed why that theory may not be plausible but still recognizing the increasing plausibility of the OBH. The purpose of this paper was not to argue in favour of the OBH as it needs further testing rather this paper was intended to show how we can use neural networks to explain some of the functions of dreams. Due to the similarities in execution, it can be concluded that neural networks can tell us a lot about brain processes which can hopefully be explored further. From everything I showed so far, dreams do play an important role in the cognition of humans. The field of marijuana research is one that, just like dreaming, still shrouded in mystery. Initial research has shown that marijuana impacts young people's mental capacities like memory but is also known that marijuana inhibits a person's REM sleep. It is my belief that there may be a link between these two facts. Maybe, marijuana or other drugs themselves do not affect a person's cognition, rather it is a side effect from the loss of REM sleep due to the brain being unable to generalize etc. This is my piece of informed speculation and possibly an avenue into using the OBH and by extension neural networks to explain much more than just dreams.

References

Andrillon, T., Windt, J., Silk, T., Drummond, S., Bellgrove, M. A., & Tsuchiya, N. (2019). Does the mind wander when the brain takes a break? Local sleep in wakefulness, attentional lapses and mind-wandering. *Frontiers in neuroscience*, *13*, 949.

Cao, J., Herman, A. B., West, G. B., Poe, G., & Savage, V. M. (2020). Unraveling why we sleep: Quantitative analysis reveals abrupt transition from neural reorganization to repair in early development. *Science advances*, *6*(38), eaba0398.

Crick, F., & Mitchison, G. (1983). The function of dream sleep. *Nature*, *304*(5922), 111-114.

Domhoff, G. W. (2003). *The scientific study of dreams: Neural networks, cognitive development, and content analysis*. American Psychological Association.

Fachechi, A., Agliari, E., & Barra, A. (2019). Dreaming neural networks: forgetting spurious memories and reinforcing pure ones. *Neural Networks*, *112*, 24-40.

Foulkes D., Domhoff G.W. (2014). Bottom-up or top-down in dream neuroscience? a top-down critique of two bottom-up studies. Consciousness and cognition, 27:168–171

Hoel, E. (2020). The Overfitted Brain: Dreams evolved to assist generalization. *arXiv preprint arXiv:2007.09560*.

Keshavan, M. S., & Sudarshan, M. (2017). Deep dreaming, aberrant salience and psychosis: connecting the dots by artificial neural networks. *Schizophrenia research*, *188*, 178-181.

King, I., Wang, J., Chan, L., & Wang, D. (2006). *Neural Information Processing: 13th International Conference, ICONIP 2006, Hong Kong, China, October 3-6, 2006. Proceedings, Part I*. Springer Berlin Heidelberg. 211-218

Meerlo, P., Benca, R. M., & Abel, T. (Eds.). (2015). *Sleep, neuronal plasticity and brain function* (Vol. 25). Springer.

Pagel J.F, Kirshtein P., (2017). Chapter Nine - Interpreting the AI Dream, Machine Dreaming and Consciousness, Academic Press. 129-135.

Rasch, B., & Born, J. (2013). About sleep's role in memory. *Physiological reviews*.

Walker, M. P., & Stickgold, R. (2006). Sleep, memory, and plasticity. *Annu. Rev. Psychol.*, *57*, 139-166.

Yuval N., Giulio T. (2010). Dreaming and the brain: from phenomenology to neurophysiology, Trends in Cognitive Sciences, Volume 14, Issue 2. 88-100.

Zhang, C., Vinyals, O., Munos, R., & Bengio, S. (2018). A study on overfitting in deep reinforcement learning. *arXiv preprint arXiv:1804.06893*.