

Towards an Deeper Understanding of the Hierarchical Temporal Memory Algorithm

DJ Passey

August 3, 2022

History of Hierarchical Temporal Memory

HTM, or Hierarchical Temporal Memory is an artificial intelligence architecture rooted in neuroscience and justified by biological plausibility. It is the outgrowth of funding by Jeff Hawkins, the former CEO of Palm Pilot and computational techniques developed by the research scientists Dileep George and Subatai Ahmad. In 2004, Jeff Hawkins published a book titled, "On Intelligence: How a New Understanding of the Brain Will Lead to the Creation of Truly Intelligent Machines". He founded the Redwood Institute for Neuroscience in 2005. The institute hired Dileep George, a computational scientist, and one year later in 2006, George gave a presentation titled "Hierarchical Temporal Memory: Theory and Applications" [5]. Three years later, in collaboration with their new company Numenta, George and Hawkins published a comprehensive theoretical paper: "Towards a Mathematical Theory of Cortical Micro-circuits" [6]. This paper describes a tree-like structure of nodes. Each node contains multiple Markov chains of which one is chosen to be "active". The state of the parent nodes is determined by which Markov chains are active in each child node. The model is optimized via Bayesian belief propagation so that probabilities of a particular Markov chain can be determined by the state of a node and it's siblings.

Dileep George left Numenta in 2010. This *YCombinator thread* suggests that George left due to a difference in technical direction. This *article* quotes George saying "HTM was an important effort, much like Poggio's foundational HMAX model, HTM had the right high level goals. But, when you dig into the algorithm level, you'll see that HTM implementations hadn't solved the problems of information representation in the hierarchy."

With the departure of George, leadership of the Numenta Research team switched to Subatai Ahmad and in 2016-2017, the company published a series of papers on their new HTM algorithm [2, 4, 3, 7].

This algorithm relies heavily on sparse distributed representations, a high dimensional method for encoding data with many useful properties [?]. The new hierarchical temporal memory algorithm divides the architecture into four components, an encoder component, a spacial pooler component, a temporal

memory component and a machine learning classifier. The first two components of the HTM architecture focus on translating raw data into an appropriate sparse high dimensional representation while the temporal memory component and classifier do the work of prediction.

This algorithm showed promise, out performing LSTM on time series prediction and showing excellent abilities at anomaly detection.

It was also a departure from the theoretical work completed by Dileep George in 2009. True to it's name, the temporal memory component no longer leverages a hierarchy of features at different granularity. Additionally, the temporal memory algorithm is optimized via a number of heuristic rules and does not rely on Bayesian optimization.

How does HTM compare to Deep Learning?

Time Series Prediction

There are a few papers compare HTM to deep learning algorithms directly [10, 9, 4, 2]. Of these, [10] is the most comprehensive. It suggests that it is the first paper to make such a comparison, and it shows definitively that with proper hyper parameter optimization, LSTM outperforms HTM, both in terms of accuracy *and speed*. It goes further than this, by using the same datasets as the original Numenta papers and showing that with proper hyper parameters, LSTM actually outperforms HTM [4, 2].

However, the future of HTM is not all bleak. The last section of [10] describes training the LSTM and HTM models built for the Numenta taxicab dataset (without changing any hyper parameters) on a non stationary dataset with a global trend. HTM was able to adapt to this dataset without any hyper-parameter adjustment. With no hyper parameter optimization, LSTM failed. While this is more of a case study, it does illustrate that HTM is fundamentally different from LSTM and offers some comparative advantages.

Anomaly Detection

Struye et. al. claims that HTM has proved itself at anomaly detection [10]. Numenta has published a *a repository* for benchmarking real time streaming anomaly detection. It is assumed that the adaptive nature of HTM makes it well suited for this sort of problem. However futher investigation calls this claim into question. This is well illustrated by the title of a recent paper: "Current Time Series Anomaly Detection Benchmarks are Flawed and are Creating the Illusion of Progress" [11]. The paper examines many anomaly detection time-series datasets including the Numenta benchmark and finds that they contain mislabeled data, many trivial problems, too many anomalies and a run to failure bias (most anomalies occur at the end). The paper uses one of the Numenta benchmark datasets as an illustration of a poor choice. The paper goes into

depth on the use of anomalies in the Numenta taxi driver dataset and shows the existence of multiple false positives and false negatives in the labels.

Data hygiene notwithstanding, HTM was beaten by another brain inspired algorithm called Adaptive Resonance on it’s own benchmark [1].

Why Study Hierarchical Temporal Memory?

The question remains, if HTM is less accurate than other techniques, what is the value of studying HTM? The answer is as follows.

While HTM technically loses to LSTM in terms of speed and accuracy, it is not far behind. This is remarkable considering that the optimization step of HTM relies on heuristic rules that are not well understood. Additionally, LSTM optimization has been supported and explored by a massive army of machine learning engineers while HTM has seen no such investment.

In our work studying the HTM algorithm, we focused on uncovering the essential mechanisms of their temporal memory algorithm: What is it *exactly* that leads to the success of this approach?

In focusing on this question we uncovered theoretical concepts that were *fundamentally different* than the theoretical tools applied by standard deep learning approaches.

Our hope is that by reducing the key mechanisms of HTM to a theoretical framework, we can leverage existing theory from sparse distributed memory, branching processes and other tools to converge on clearer optimization pathways, lighter weight models and the potential to augment current machine learning techniques with fresh tools. Additionally, HTM appears to have a comparative advantage when it comes to adaptability and resistance to certain types of noise [2, 3, 4]. It is possible that a deeper investigation of the relevant computational mechanisms will provide a pathway to refine this comparative advantage.

References

- [1] Leonardo Enzo Brito da Silva, Islam Elnabarawy, and Donald C. Wunsch. A survey of adaptive resonance theory neural network models for engineering applications. *Neural Networks*, 120:167–203, 2019. special Issue in Honor of the 80th Birthday of Stephen Grossberg.
- [2] Yuwei Cui, Subutai Ahmad, and Jeff Hawkins. Continuous online sequence learning with an unsupervised neural network model. *Neural Computation*, 28(11):2474–2504, nov 2016.
- [3] Yuwei Cui, Subutai Ahmad, and Jeff Hawkins. The htm spatial pooler—a neocortical algorithm for online sparse distributed coding. *Frontiers in Computational Neuroscience*, 11, 2017.

- [4] Yuwei Cui, Chetan Surpur, Subutai Ahmad, and Jeff Hawkins. A comparative study of htm and other neural network models for online sequence learning with streaming data. In *2016 International Joint Conference on Neural Networks*, pages 1530–1538, 07 2016.
- [5] Dileep George. Hierarchical temporal memory: Theory and applications. In *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, volume Supplement, pages 6643–6643, 2006.
- [6] Dileep George and Jeff Hawkins. Towards a mathematical theory of cortical micro-circuits. *PLoS computational biology*, 5:e1000532, 10 2009.
- [7] Jeff Hawkins and Subutai Ahmad. Why neurons have thousands of synapses, a theory of sequence memory in neocortex. *Frontiers in Neural Circuits*, 10, 2016.
- [8] Wolfgang Maass, Christos H. Papadimitriou, Santosh Vempala, and Robert Legenstein. *Brain Computation: A Computer Science Perspective*, pages 184–199. Springer International Publishing, Cham, 2019.
- [9] Jonathan Mackenzie, John F. Roddick, and Rocco Zito. An evaluation of htm and lstm for short-term arterial traffic flow prediction. *IEEE Transactions on Intelligent Transportation Systems*, 20(5):1847–1857, 2019.
- [10] Jakob Struye and Steven Latré. Hierarchical temporal memory and recurrent neural networks for time series prediction: An empirical validation and reduction to multilayer perceptrons. *Neurocomputing*, 396:291–301, 2020.
- [11] Renjie Wu and Eamonn Keogh. Current time series anomaly detection benchmarks are flawed and are creating the illusion of progress. *IEEE Transactions on Knowledge and Data Engineering*, pages 1–1, 2021.