Interpretation of "Neural Network as the World"

Stephane H. Maes[1]

September 14, 2020

Abstract:

Recently a controversial series of paper ends up proposing the possibility that the universe be a neural network by observing that with an irreversible thermodynamics model of the learning process of the neural network, it might appear possible to model quantum and quantum physics as well as observe emergence of a General Relativity space time and gravity, and the plausibility to construct a generalized holographic principle beyond the AdS/CFT correspondence conjecture. The approach has been received with some skepticism.

In this paper[2], we present our suggestions to interpret the results including how NN models could relate to the Wigner wonder at why mathematics describe the Physical world.

____

1. Introduction

[2] shows that if information theory is modeled with (covariant) irreversible / non-equilibrium thermodynamic processes then, close to equilibrium, the conjugate thermodynamics variables of the information content (tensor) is an emerging spacetime following the Hilbert Einstein spacetime. This result is to be related to [4], that derives emergence of quantum mechanics from classical irreversible thermodynamics. Away from equilibrium, the picture is less clear. We note that the irreversibility has to be directly related to the quantum behavior.

Following on these results, [3] proposes a thermodynamics model for Machine Learning (ML) and derives a proposal for Thermodynamics of learning. NN are example of ML, but we know that any AI or ML algorithm can always be modeled as a NN [12-14].

[1] then models NN thermodynamics, using [3] and inspiring from [2,4] and shows:

- Close to equilibrium and when the entropy contributions from learning are small, one can recover a Schrödinger equation and a wave function that results from the stochastic dynamics of the training variables randomly trying to find where to go to learn. It amounts to small scale events, trying different evolution to find hints of the best ones, not really changing much with respect to what the NN has learned, and it denotes a state of the NN, where equilibrium has been reached, and new variables values for the models are randomly visited just in case they could help or because learning continues.
- Further away from equilibrium, where random fluctuations of the q_i, the learning variables, are smaller and less visible, and hence at larger scales, and therefore when learning process dominates the thermodynamic, the training variable have an evolution that can be characterized by a classical

[1] shmaes.physics@gmail.com

[2] This is a republication of an appendix in [5], presented independently of notions of multi-fold universe that are not needed for the present analysis.

Hamiltonian and therefore can be modeled by classical Physics. It corresponds to a state of the NN, where it can estimate how to progress to learn or improve the loss/cost function (think of gradient like steepest gradient descent methods for learning/training/optimization).

- When modeling directly the dynamics of the state of the neurons, [2] applies and under suitable conditions (close to equilibrium and with weak interaction between the neurons (at least when nonlocal)), the dynamics of the neurons follows Einstein's GR field equations.
- Analyzing In and Out layers of the NN versus hidden layers, one can hypothesize ways to recover a generalized holographic principle that would link a quantum mechanically dominated NN (In + Out layers) to a deep / many layers NN dominated by gravity.

Section 2 presents additional considerations on what we can learn from [1].

However, this model does not model entanglement yet. It is a key missing part before we can claim to have a truly complete quantum model emerging from [1]. A proposal to that effects is presented in [5] and it is interpreted as hinting the multi-fold mechanisms proposed in [11].

2. A NN model of the world? An alternate interpretation

[1] proposes that the universe is a NN. We do not believe that this is the only interpretation of the results presented in [1] and we want to propose an alternative explanation. As already mentioned, the NN approach can be seen as a model of the dynamics of Physics in the universe. Such model is mathematical, in fact it is a consequence of Hilbert 13th problem and the ability to model any system with deep hidden layers and in particular NN as demonstrated with the Kolmogorov-Arnold representation Theorem [6] and the Universal approximation theorem [7].

In the present case the dynamics of the state variables, i.e. the equation of motion, are the approximated functions. Per the theorems above we know such approximation is (almost) always possible (up to discontinuities) and to any desired degree of accuracy (for the right optimization strategy in the case of NN).

What is interesting, is that if the algorithm for loss/cost function optimization relies on (classical) Thermodynamics (for Irreversible and for non-equilibrium processes with a Free Energy model), it uncovers naturally the dynamics described in section 1 and in [1], where the fact that the NN includes also the model of the learning processes allows to capture in one shot dynamics of the physical system (i.e. the universe) and the dynamic of information processing; therefore concretizing the physical information theory aspects also (e.g. see [8] for related aspects of physical information theory); something that now can be captured into a common Thermodynamics (and physical) model. It goes beyond [4] and justifies considerations like Learning's Thermodynamics or the principle of conservation of information. In our view, much more than having a NN modeling (or being per [1 the universe, the key aspect is that we have a complete model for physical and information entropy modeling and computing.

In such a model, it makes sense that entropy extremization and action extremization become equivalent or dual. It is also natural to see that, at small scales, quantum fluctuations around equilibrium imply fluctuations of the learning variables, and the NN state, while at larger scales away from equilibrium (albeit still close), the system will rather behave classically as a learning system (to go back to equilibrium).

So we interpret [1] as a model that shows first and foremost how Physics + Information Theory coexist into a larger model. The model of [1] has its own dynamics. These dynamics may be seen as a model of how physical systems like the universe handle information conservation or just as an algorithm to derive the same outcome. More work is needed to determine that. If it is the former, this may actually be a way to answer why and how mathematics are so good at modeling the Universe as asked famously by Wigner [20], and others, and it would be aligned with Tegmark's view [10]. Indeed, [1] would now amount to modeling how the universe remains close to

thermodynamic equilibrium while always reacting to changes and fluctuation (e.g. random, thermal external, etc.) to catch up with the mathematical prescription aiming at optimizing the loss/cost function while evolving with minimum disruptions as captured by extremization of the entropy and action changes: physical systems take some "guessed optimized efforts" to catch up and follow the mathematics that describe them correctly and these mathematics are the reflection of this process. It is a direct application of Pontryagin's maximum principles and theorem [25-27].

3. Conclusions

There have been already many hints of relationships between spacetime, entanglement, thermodynamics and information theory like treating the universe as universal Quantum Computer, encountering error correcting code in spacetime (including in [11]), deriving GR from spacetime properties in equilibrium and the relationships between gravity, entropy and entanglement entropy as well as the principle of conservation of information in Quantum Physics and the information paradox with Black holes. Information and Physics are closely related and this paper, along with many of its references, add to these observations.

We discussed how [1] can be understood intuitively as correctly modeling the universe in particular as a universal quantum computer and how its result may relate to Wigner's question about why mathematics describe so well the physical world.

We refer to [5] for discussion on how to add entanglement to the model.

———

References:

[1]: Vitaly Vanchurin, (2018), " The world as a neural network", arXiv:2008.01540v1
[2]: Vitaly Vanchurin, (2018), "Covariant Information Theory and Emergent Gravity", arXiv:1707.05004v4
[3]: Vitaly Vanchurin, (2020), "Towards a theory of machine learning", arXiv:2004.09280v3
[4]: D. Acosta, P. Fernandez de Cordoba, J. M. Isidro, J. L. G. Santander, (2012), "Emergent quantum mechanics as a classical, irreversible thermodynamics", arXiv:1206.4941v2
[5]: Stephane H Maes, (2020), "Implicit Multi-Fold Mechanisms in a Neural Network Model of the Universe", viXra:2012.0191v1, https://shmaesphysics.wordpress.com/2020/09/12/implicit-multi-fold-mechanisms-in-a-neural-network-model-of-the-universe/, September 12, 2020.
[6]: Wikipedia, "Kolmogorov–Arnold representation theorem" https://en.wikipedia.org/wiki/Kolmogorov%E2%80%93Arnold_representation_theorem, Retrieved on September 14, 2020.
[7]: Wikipedia, "Universal approximation theorem", https://en.wikipedia.org/wiki/Universal_approximation_theorem, Retrieved on September 14, 2020.
[8]: Seth Lloyd, (2006), "Programming the Universe: A Quantum Computer Scientist Takes on the Cosmos", Alfred A. Knopf
[9]: Wigner, E. P. (1960). "The unreasonable effectiveness of mathematics in the natural sciences. Richard Courant lecture in mathematical sciences delivered at New York University, May 11, 1959". Communications on Pure and Applied Mathematics. 13: 1–14.
[10]: Max Tegmark, (2007), "The Mathematical Universe", arXiv:0704.0646v2
[11]: Stephane H. Maes, (2020), "Quantum Gravity Emergence from Entanglement in a Multi-Fold Universe", viXra:2006.0088v1, https://vixra.org/pdf/2006.0088v1.pdf (June 9, 2020).
[12]: Wikipedia, "Kolmogorov–Arnold representation theorem" https://en.wikipedia.org/wiki/Kolmogorov%E2%80%93Arnold_representation_theorem, Retrieved on September 14, 2020.
[13]: Wikipedia, "Universal approximation theorem", https://en.wikipedia.org/wiki/Universal_approximation_theorem, Retrieved on September 14, 2020.

(Added when pre-print was published on vixra.org)

[14]: Andre Ye, (2020), "Every Machine Learning Algorithm Can Be Represented as a Neural Network", https://towardsdatascience.com/every-machine-learning-algorithm-can-be-represented-as-a-neural-network-82dcdfb627e3. Retrieved on December 19, 2020

[15]: Wikipedia, "Pontryagin's maximum principle", https://en.wikipedia.org/wiki/Pontryagin%27s_maximum_principle**.** Retrieved on September 29, 2020.

[16]: "13 Pontryagin's Maximum Principle", http://www.statslab.cam.ac.uk/~rrw1/oc/L13.pdf. Retrieved on September 29, 2020.

[17]: Thayer Watkins, "The Nature of the Principle of Least Action in Mechanics", https://www.sjsu.edu/faculty/watkins/minprin.htm. Retrieved on September 29, 2020.