WHAT DOES IT MEAN TO UNDERSTAND? NEURAL NETWORKS CASE

A TRACTATUS

Albert Ierusalem
albertierusalem@protonmail.com

Aleksandr Senin
aleksandr.arxstuff@yandex.ru

October 28, 2019

ABSTRACT

We can say that we understand neural networks then and only then if you will come to me and say that the best model ever for some task has a 100 layers, and I will answer "No! 101 layers model is the best!".

***K*eywords** Understanding · Neural Networks

1 Motivation

1.00. In this paper, we express our opinion on the neural network's understanding issue. We were motivated by Timothy P. Lillicrap and Konrad P. Kording paper[0] and we disagree with many statements presented in this work. We propose understanding requirements and based on them, describes the state when we can say that we understand the neural networks.

2 Background

2.00. Rational: structured, holistic, impassive, abstract.

2.01. Irrational: disjointed, preconceived, true.

2.02. Rational is "consistency", the irrational is "trust".

2.03. The irrational is indivisible: the division of the irrational is "distrust", rational is divisible and consistent.

2.04. Emergence = unexpectedly emergence.

2.05. The more emergence in the system, the more heuristics it requires to deal with it.

2.06. Description produce understanding, understanding produce knowledge.

2.07. Knowledge destroys heuristics.

2.08. The knowledge is twofold in its essence: it is rational and irrational.

2.09. To gain knowledge, understanding of the rational and irrational essence of phenomena is necessary.

2.10. Understanding exclusively rational part leads to confusion, irrational - to mediocrity.

2.11. The irrational describes the texture, the rational describes the structure.

2.12. Rational is fundamental, irrational is mercurial.

2.13. The more superficial knowledge of the phenomena, the more irrational it is, and the easier it is to make mistakes.

2.14. The deeper the knowledge about the phenomena, the more rational it is and the more difficult to use it.

2.15. Important to understand how much we understood something if it not fully understood yet.

3 The Requirements

3.00. System consist of processes.

3.01. Process = phenomena = fact.

3.02. A simple fact is an indivisible fact.

3.03. A simple fact is always understood.

3.04. A complex fact is a divisible fact, there is no upper bound.

3.05. For a simple fact, all the possibilities of its manifestation in a complex fact are known.

3.06. The existence of complex facts is predetermined in a simple fact, and the existence of a simple fact in a complex is predetermined in a complex.

3.07. All real systems are dynamic.

3.08. A dynamic system produces new facts over time.

3.09. A description is a specification of the causal relationship between a set of system facts noted as simple facts.

3.10. All descriptions are constant.

3.11. In the description, the complexity of the facts is blurred and does not matter.

3.12. One complex system can be described in different abstract levels.

3.13. If we can describe some set of facts, these facts are in the same abstract level.

3.14. The set of facts described in the different abstract levels can intersect.

3.15. A description is true and produce understanding if it satisfies understanding requirements.

3.16. First requirement: system described in the all possibly abstract levels. (We call it descriptive completeness).

3.17. Second requirement: description of the system of the abstract descriptions reduces emergence to zero. (We call it descriptive emergence).

3.18. The difficulty of understanding is in finding all abstract levels of the system.

3.19. If after describing the system by one or more abstract levels, emergence persists in the system, it is necessary to get more abstract level descriptions, to bridge the gap.

3.20. The greater the gap between abstract description levels, the bigger the emergence between facts of these levels.

3.21. Third requirement: proposed description does not contradict to the new coming facts. (We call it descriptive power)

3.22. We cannot evaluate the description if we have no facts beyond the scope of this description

3.23. If a satisfied lower level requirement does not lead to satisfying a higher requirement, this description is not true.

3.24. Satisfying at the higher level guarantee satisfying at lower levels.

3.25. Descriptive stability, completeness, and power are not binary values, they are continuous, from low to high.

3.26. With the same completeness, power, and emergence, the compact description is preferable.

3.27. Many systems describe by an insufficient number of abstract levels, thus, the emergence easily arises.

3.28. Often the understanding it's a fiction.

3.29. Scientific research of any system will be completed after description satisfying all understanding requirements.

4 Understanding Neural Networks

4.1 Lillicrap et al. statement

5.100. We can define a neural network that can learn to recognize objects in less than 100 lines of code. However, after training, it is characterized by millions of weights that contain the knowledge about many object types across visual scenes. Such networks are thus dramatically easier to understand in terms of the code that makes them than the resulting properties, such as tuning or connections.

5.101. Maybe humans or other intelligent beings can figure out ways of meaningfully arguing about non-compact systems, but, at least so far, compactness is necessary for what we would call a meaningful understanding...

5.102. There is a widespread belief in neuroscience that there can be a meaningful mid-level model of the dynamics of the brain that both can be communicated and work for complex tasks. The idea is in analogy to physics: In quantum mechanics, the state of e.g. a gas can not be meaningfully communicated since every gas atom may exist in a high dimensional space and there are many atoms. But in statistical physics, the state of the gas may be compactly described in terms of a few variables like temperature, pressure, etc. And these mid-level descriptions can afford high accuracy predictions and control. From this analogy it is sometimes argued that a compact midlevel description should exist for neuroscience.

5.103. We concluded above that for artificial neural networks a focus on objectives, learning rules, and architectures is most promising. We do not see why the same argument does not carry over to neuroscience...

4.2 Answer

5.200. To understand the code of neural network, doesn't mean to understand the neural network.

5.201. We can't say that we understand neural networks describing them using program code as we can't say that we understood the brain, describing it using physics.

5.202. What if we have 1000 weights and 10,000 lines of code for the model? We still understand the code better, even if it not compact.

5.203. We understand lines of code better because it fully satisfying understanding requirements.

5.204. All abstract levels of code are known. We have a zero emergence between different abstract levels of program code, from low to high.

5.205. If we made a mistake in the code, we can always describe how these mistakes affect the other processes, and why it's a mistake, there is no emergence in the system.

5.206. If we made a mistake in deep neural network architecture, we can't always describe how these mistakes affect the other processes, and why it's a mistake.

5.207. Program code facts are only one of the all possibly abstract level descriptions for the facts produced by the neural network.

5.208. Description based on such a set of abstract levels does not satisfy the second requirement.

5.209. Description based on program code of objectives, learning rules, and architectures is full of emergence for the neural networks facts.

5.210. Using such a code base description on practice we searching for some heuristics to make neural network works.

5.211. Approach with such random heuristics search in neuroscience seems at least strange.

5.212. The quality of control and the number of gas facts that described using statistical physics is so low, that it proposes to have such a description for neuroscience similar to the suggestion not to study neural networks at all.

5.213. Statistical physics gas description is only one of possibly abstract descriptions.

5.214. Statistical physics gas description is an example of understanding that produced by a single abstract level description and has low completeness, high emergence, and low power.

5   Conclusion

6.00. To understand the neural networks mean to have a description that satisfies all requirements.

6.01. To understand the neural networks mean to have a description that: describes the system on all abstract levels, a causal relation between which produces a low emergence between simple and complex facts, and not contracting to the new facts produced by the system.

6.02. To understand the neural networks mean to have a description that:

Define the facts produced by neural networks, for example, the activation function is a fact, dynamical isometry of the network is a fact, dynamical isometry state under some activation is a fact.

Describe facts on the different abstract levels. Abstract levels are something soft, there is no strong rule on how to define facts into some abstract level (it's research).

Reduced emergence between simple and complex facts to zero, for example: based on some trained network properties, we can predict the performance of the network on a given task.

Not contracting to the new facts produced by the system.

6.03. Stop pretending that artificial neural networks are already such architecture of knowledge from which we can transfer to the real neural network.

For artificial neural networks, we need a description that allows us to find the best model for a given task with a minimum heuristics. We need to understand it, or rebuild it, to know how to expand them to strong intelligence.

References

[0] Timothy P. Lillicrap, Konrad P. Kording. What does it mean to understand a neural network? URL:https://arxiv.org/pdf/1907.06374.pdf