Understanding Foundation of Neural Network: Reasoning-based Approach

Jihyeon Yoon*
(Department of Physiology, College of Korean Medicine, Daejeon University)

November 27, 2022

Abstract

Background: Currently, neural network model needs understanding on what derives conclusion by execution. Neural network model imitates human intelligence but it has a limitation of cannot explaining what it is to construct the knowledge which it derives.

Object: By inducing universal framework from existing neural network models, common reasoning that regulates the intelligence that derives from the models would be introduced.

Method: Starting from minimizing model's bias, integrative method of understanding how neural network takes place in numerical ordinals and cardinals would be introduced. After that, explanation of intelligence by neural network would be described.

Result: Foundation of artificial neural network intelligence is an optimization with a grid of minimized error from ground truth.

1 Introduction

Current understanding of neural network is limited by their yet-to-be explained portion–which is commonly called as "the black box-like attribute of the neural network".

This black-box likeness mostly induces disinformation of environment of a model exists.

1. As phenomenon resembles of a computer imitating an answer without any knowledgeable reasoning, a computer itself seems to take a higher ground of the creator.
2. Likelihood of fitting answer imitating the reality makes illusion of getting right answer being not about getting best answer. So-called ad-hoc correspondence in getting solution for a situation momentarily anesthetize the pain but doesn't solve the kernel problem.
3. Cost to fill a gap between best answer and ad-hoc answer occurs moment by moment and creates ambiguous bias more than it used to be before the moment.

Consequently, these cause imputation on a model and its creators. And the bias–which the original intention of model targeted to minimize–increases.

Then it's human's obligation to eliminate the black box portion of intelligence imitating model in order to recover from preceding discussion. If this "black box" portion could be explained appropriately, it could be said that the method by reasoning with these mathematical/statistical models could be safe to be used by human.

2 Review of Previous Studies

There could be two approaches for perceiving models. One is by input by singular or sequence. And the other is by output by singular or sequence. As sequence has a premise of being described in time, it could be said that by space(singular) and time(sequence), input and output of a model described.

Then, it could be said that there are four kinds of model by upper discussion.

- Model Type 1 is mainly about regression between given numbers to another space. Multiple cardinals regress into a form of model.

*Jihyeon Yoon is a korean medicine doctor. And he is a freelance programmer also.
E-mail: flyingtext@edu.dju.ac.kr
Yeongdeungpo-gu, Seoul, 07429,
Seoul, South Korea
ORCiD: https://orcid.org/0000-0001-9610-0994

		Input	
		Single Input (Space-orientated)	Sequential Input (Time-orientated)
Output	Single Output (Space-orientated)	Model Type 1 $\Rightarrow$ Cardinal(s) to Cardinal(s)	Model Type 2 $\Rightarrow$ Ordinal(s) to Cardinal(s)
	Sequential Output (Time-orientated)	Model Type 3 $\Rightarrow$ Cardinal(s) to Ordinal(s)	Model Type 4 $\Rightarrow$ Ordinal(s) to Ordinal(s)

Table 1: Model by Input and Output Type

- In Model Type 2, if ordinality expands to a corresponding cardinal amount, it resembles Gaussian distribution. Clustering could be addressed in this model.
- In Model Type 3, if cardinality constitutes an ordinal sequence, it is based on χ^2 distribution. Classification could be addressed in this model.
- Model Type 4 is mainly about generating space from another space. This could be done by assembling multiple models with upper Model 1, 2, and 3.

Former neural network studies focused on how each distribution could be gained by imitating existing examples. But there comes question when facing numerical model constructs itself. What does the number that constructs the model means specifically? What is the specific meaning of the given matrix of graph when it could be applied in specific problem solving method? And most importantly, How could the matrix that shows relations imitate the reality?

3 Optimization for Information Processing

Any kind of Neural Network is based on the output of numerical distribution which could be related onto regression or classification. School of pedagogy looks at human learning in three different ways. Likewise, in this paper, the principles of building neural networks based on three methodologies would be explored.

3.1 Construction as a belief

Constructing neural network as a belief is about minimizing doubt. It's mainly based on Constructivism on behalf of constructing model by their original purpose. How is a neural network configured? Most intuitively, it is constructed by chain product of matrices and sum of the preset constants. What does this mean specifically? There can be two perspectives.

$$\begin{aligned}
&Assuming\ that,\\
&Input\ \vec{I},\ Output\ \vec{O},\ n\text{–}th\ Layer\ Matrix\ L_n,\ n\text{–}th\ Bias\ term\ B_n, Activation\ Function\ A\\
&Neural\ Network\ Model\ M \models\ \vec{O} = ...A\left(A\left(A\left(A\left(\vec{I}\cdot L_1 + B_1\right)\cdot L_2 + B_2\right)\cdot L_3 + B_3\right)\cdot L_4 + B_4\right)...
\end{aligned}$$

First perspective is on the matrices. Chaining product of matrices by being activated or not implies that it multiplies original vector value to cast the relation to consecutive layers. As for even a single number could be decomposed into multiple matrices, relation by graphs could be set and controlled by activation function.

Second perspective is on the biases. Controlling amount of center value while chaining matrices implies that each chaining layer of the vectors in a matrix is controlled to point distributional similarity of the training data. Intuitively, it's transformation of vectors by moving from original point before chaining to consecutive layer to transform.

Then the each layer of neural network process could be in process:

1. Transform input vectors $\Longrightarrow \vec{I_{i-1}}\cdot L_i$
2. Move input vectors $\Longrightarrow \vec{I_{i-1}}\cdot L_i + B_i$
3. Check activation $-1 \leq A(\vec{I_{i-1}}\cdot L_i + B_i) \leq 1$
4. Sum up by activation as matrix grid $\Longrightarrow \vec{I_i} = \sum_j A(\vec{I_{i-1}}\cdot L_i + B_i)_j$

Layer 1 (%)	More Bias	Less Bias
More activation	A	B
Less activation	C	D

Table 2: Layer 1 Confusion Matrix

Layer 2 (%)	More Bias	Less Bias
More activation	E	F
Less activation	G	H

Table 3: Layer 2 Confusion Matrix

Layer 1 (%)	More Bias			Less Bias		
More activation	Layer 2 (%)	More Bias	Less Bias	Layer 2 (%)	More Bias	Less Bias
	More activation	A$\cdot E$	A$\cdot F$	More activation	B$\cdot E$	B$\cdot F$
	Less activation	A$\cdot G$	A$\cdot H$	Less activation	B$\cdot G$	B$\cdot H$
Less activation	Layer 2 (%)	More Bias	Less Bias	Layer 2 (%)	More Bias	Less Bias
	More activation	C$\cdot E$	C$\cdot F$	More activation	D$\cdot E$	D$\cdot F$
	Less activation	C$\cdot G$	C$\cdot H$	Less activation	D$\cdot G$	D$\cdot H$

Table 4: Layer 1 & 2 Confusion Matrix (and so on)

This could be applied to understand confusion matrix(Table 4). Consequently, more the layers, more division by model occurs. In the interim conclusion, the weights and numbers in the deep learning graph model are composed of operational vectors that cause relative changes to the input vectors in layer before.

3.2 Function as a model

Functioning neural network as a model is about minimizing unexpected variable. It's mainly based on functionalism on behalf of given data to be fit into a model. Then, how could this achieved?

3.2.1 Meaning of error notation in neural network

Then how can the error of neural network be written? The error notation of the neural network is the same as the cardinal notation of the distance from result to the objective of a model. (Like L_1, L_2, and etc.) And within previous discussion, neural network's goal notation is bound to be determined by ground truth by input dataset. Likewise, the meaning and purpose of a model is distinct to the construction of a model. This implies two things.

- Relationship between complexity and ethics of a model - ethical compliance for usage of a model, from simple to complex models, must be applied consistently.
- Relationship between complexity and superiority of a model - complexity of the model does not guarantee the superiority of the model in usage of a model.

3.2.2 Reducing error by training in neural network

How can neural network reduce errors? It is commonly known as training of neural networks by ground truth and back propagation of errors. And the error of the neural net is indicated by the distance from the ground truth.

In other words, the activation weight value of the neural network is the error notation of the result value and the ground truth, and the bias value of the neural network corresponds to the original difference between the result value and the ground truth. Therefore, non-errors (weights, biases) in neural network corresponds to the multiplication of the values leading to vector transformations representing ground truth.

Upper discussion conceived neural network model as a chain product of layers and their sum of bias.

$$
\begin{aligned}
&Assuming\ that\ Matrix\ \forall m_x\,(x \leq n(L))\,,\\
&Neural\ Network\ Model\ M \models\ \vec{O} = ...A\left(A\left(A\left(A\left(\vec{I}\cdot L_1 + B_1\right)\cdot L_2 + B_2\right)\cdot L_3 + B_3\right)\cdot L_4 + B_4\right)...\\
=&\vec{I}\cdot(m_1\cdot m_2\cdot m_3\cdot m_4\cdots m_x)\cdot B_1 + \vec{I}\cdot(m_2\cdot m_3\cdot m_4\cdots m_x)\cdot B_2 + \vec{I}\cdot(m_3\cdot m_4\cdots m_x)\cdot B_3\cdots\\
&+\vec{I}\cdot m_x\cdot B_x
\end{aligned}
$$

And from preceding discussion [1], three types of bias that could occur have been introduced.

$$\begin{aligned}
&Assuming\ that,\\
&PA = Partial(Batch)\ Activation\\
&TA = Total\ Activation\\
&PAD = Partial(Batch)\ Activation\ Distribution
\end{aligned}$$

$$\begin{aligned}
&Opportunity\ fallacy\ probability\ function\ OE(M) =\\
&\quad Pr(\{M_{oe} \mid M_{oe}\ is\ a\ model\ when\ \frac{observed\ PA}{observed\ TA} > expected\ PAD\} \mid \forall M)\\
&Memory\ fallacy\ probability\ function\ ME(M) =\\
&\quad Pr(\{M_{me} \mid M_{me}\ is\ an\ model\ when\ \frac{observed\ PA}{observed\ TA} < expected\ PAD\} \mid \forall M)\\
&Perception\ fallacy\ probability\ function\ PE(M) =\\
&\quad Pr(\{M_{pe} \mid M_{pe}\ is\ an\ model\ when\ \frac{expected\ PA}{expected\ TA} \neq (observed\ PAD\} \mid \forall M)
\end{aligned}$$

- Opportunity error probability is probability of an error which is an underestimation in executed activation than reality.
- Memory error probability is probability of an error which is an overestimation in executed activation than reality.
- Perception error probability is probability of an error which is an distribution of batch set doesn't match ground truth.

Continuing upper discussion, $B_1, B_2, B_3 \cdots B_n$ could be expressed like this:

$$\begin{aligned}
&Neural\ Network\ Model\ M \models\ \vec{O} = ...A\left(A\left(A\left(A\left(\vec{I}\cdot L_1 + B_1\right)\cdot L_2 + B_2\right)\cdot L_3 + B_3\right)\cdot L_4 + B_4\right)...\\
=&\vec{I}\cdot(m_1\cdot m_2\cdot m_3\cdot m_4\cdots m_x)\cdot B_1 + \vec{I}\cdot(m_2\cdot m_3\cdot m_4\cdots m_x)\cdot B_2 + \vec{I}\cdot(m_3\cdot m_4\cdots m_x)\cdot B_3\cdots\\
&+\vec{I}\cdot m_x\cdot B_x
\end{aligned}$$

By as an attribute of "Construction as a belief" and "Function as a model", optimized training and model compression method for neural network could be derived. Then how this could be approach to minimizing unexpected variable? It's because by the process in maximizing correction in given ground truth Correction function could be expressed like this:

$$\begin{aligned}
&\text{When Training Layer (By Definition)},\\
&Layer\ Output\ LO_n = A(LO_{n-1}\cdot L_n + B_n)\\
&Layer\ L_n = \underset{m_n}{\arg\max}\ ||\ LO_{n-1}\cdot(1-OE(m_n))\cdot(1-ME(m_n))\cdot(1-PE(m_n))\ ||\\
&Bias\ B_n = \underset{B_n}{\arg\min}\ ||\ LO_{n-1} - LO_{n-1}\cdot(1-OE(m_n))\cdot(1-ME(m_n))\cdot(1-PE(m_n))\ ||\\
&\Rightarrow\ Optimized\ M^* = \underset{M}{\arg\max}\left(\prod_x Correction(m_x) \leq 1\right)
\end{aligned}$$

3.3 Conduction as an example

Conducting neural network as an example is about minimizing social error. It's mainly based on behaviorism on behalf of adaptation to society. This is a subject for a ground truth. More specifically, this is why GAN(Generative Adversarial Networks) model is more adaptive than other existing solutions.

In preceding discussion, Table 1 presented Model Type 4, which Sequential Input to Sequential Output is a combination of ordinality and cardinality in both input and output. On the other hand, as discussed, ordinality could be expanded into a distribution of cardinality. And by that mean even a single number can be decomposed into a $(1 \times n)\cdot(n \times 1)$ matrix ($n \in N$). Conversely, if the former cardinal distribution that constitutes single number could be regulated by ordinality by true or false in another model, it would cause elaboration to generate more ground truth-induced imitation result.

State-of-the-art neural network model demonstrates many of neural network constructed as GAN(Generative Adversarial Networks) performs better than the former predecessor neural network models. This is because not only cardinal amount of error intervenes the model to structure itself, but also the check for validity that makes gap between original and the other seeks its most identifiable construction as a logic for discriminating ground truth and generation for the copy from original ground truth.

3.3.1 Discrimination as a belief

How could a discrimination between original and copy be a model that standardize the generation? It could be explained through the variability that neural network could provide. As inference from discussed earlier [1]:

$$\begin{aligned}
&\text{Assuming that, SR and CR meets conservation rule of symmetry,}\\
&Optimized\ M^* = \arg\max_M \left(\prod_x Correction(m_x) \right)\\
&= \arg\max_M \prod_x (Correction(m_x) + SR_x - CR_x)\\
&(0 \leq SR_x \leq 1,\ 0 \leq CR_x \leq 1)
\end{aligned}$$

If ground truth itself is given as a definite constant, $SR_x - CR_x$ must be 0. But if ground truth itself fulfills gauge symmetric attribute, $SR_x - CR_x$ not has to be 0. This leads to the better chance of minimizing error to the bias.

3.3.2 Generation as a model

Following upper discussion, it could be said that generation of a model from an example is better than directly giving a right answer. By symmetrical gauge attribute and its asymmetry conservation, generating copy of operation to the input data could be established. And it is related to expansion of ordinal attribute (original or copy) to the cardinal amount of error margin that feedback to the generator model. This could be a case of an education to the neural network by an example in themselves, which expands original model with the minimal data as a origin to the larger model.

$$\begin{aligned}
&Optimized\ M^* = \arg\max_M \left(\prod_x Correction(m_x) \right)\\
&= \arg\max_M \prod_x (Correction(m_x) + SR_x - CR_x)\\
&\implies Optimized\ M'^* = \arg\max_{M'} \prod_x (Correction(m_x) + Correction(m'_x) + SR_x - CR_x)\\
&\Rightarrow M \approx M'
\end{aligned}$$

4 Conclusion

1. Foundation of artificial neural network intelligence is an optimization with a grid of minimized error from standard.
2. Neural network model is consists with a bundle of correction function from given error to handle bias, which could be understood as a result of an optimization.
3. Giving an example to optimizing neural network model gives better opportunity to fit in ground truth in a sense of gauge conservation.

References

[1] J. Yoon, “Intuitive Understanding of Observing Reality: With Extension of the Game Theory,” 8 2022. [Online]. Available: https://www.techrxiv.org/articles/preprint/Intuitive_Understanding_of_Observing_Reality_With_Extension_of_the_Game_Theory/20485623