Determining the Neural Network Topology from the Viewpoint of Kuhn's Philosophy and Popper's Philosophy

1st Muh. Ibnu Choldun Rachmatullah
School of Electrical Engineering and Informatics
Institut Teknologi Bandung
Bandung, Indonesia
ibnucholdun@poltekpos.ac.id

2nd Kridanto Surendro
School of Electrical Engineering and Informatics
Institut Teknologi Bandung
Bandung, Indonesia
endro@informatika.org

3rd Judhi Santoso
School of Electrical Engineering and Informatics
Institut Teknologi Bandung
Bandung, Indonesia
judhi@stei.itb.ac.id

4th Dimitri Mahayana
School of Electrical Engineering and Informatics
Institut Teknologi Bandung
Bandung, Indonesia
dimitri@lskk.ee.itb.ac.id

***Abstract*—Determining the number of hidden layers and the number of neurons are very important and have a large influence on deep neural network(DNN) performance. In some studies, there is no clear guideline on how to determine the number of hidden layers or neurons optimally; even the roles and functions of both are explained minimally. Although it is difficult, researches to determine the number of hidden layers and neurons must continue to be carried out, because both will greatly determine the performance of DNN. According to Kuhn, the method for determining neural network topology in deciding the number of hidden layers and hidden neurons is still in pre-paradigm phase. New studies continue to be made in an effort to find methods that can be generally accepted, so that they will become normal sciences. The proposed new methods can be tested by using Popper's falsification which will determine whether the methods can eventually become normal sciences or not.**

Keywords—topology, neural network, pre-paradigm, falsification

I. INTRODUCTION

Since the increase in the computer capabilities, the use of a neural network that has more than one hidden layer has attracted interest from researchers, especially since the use of deep neural networks to solve problems in the real world. Deep neural network can be interpreted as a technique that uses neural networks for learning that utilize many hidden layers between input and output layers [1]. One of the challenges in the successful implementation of deep neural networks is setting values for various hyper parameters, one of which is the network topology, which is closely related to the number of hidden layers and neurons. Determining the number of hidden layers and neurons is very important and has a large influence on the performance of deep neural networks [2]. Determining these two manually (usually through 'trial and error' method) to find a fairly optimal topology is a time-consuming process.

Some studies on neural network topology have focused on determining the numbers of neurons because they only use one hidden layer, some focus on the number of hidden layers (one or two hidden layers), some also determine the number of hidden layers and neurons in each hidden layer. Researches on determining the number of hidden neurons have been going on since the 1990s and are still an interesting topic for researchers [3][4]. Studies comparing the performance of one or two hidden layers are still an interesting topic to date [5][6]. While researches calculating the number of hidden layers and neurons have been done in recent years, since the emergence of deep learning [1][2][5]. Determining the right number of neurons is important to avoid under-fitting or over-fitting, and is also prominent in increasing the level of accuracy of the neural network. Deciding the right number of hidden layers and neurons is important to reduce the complexity of processing time and to maintain the accuracy of the neural networks [2].

A number of methods have been carried out to calculate the number of hidden layers and neurons, for example: model-based automatic method using particle swarm optimization(PSO) [1][7], automatic method without model using grid search (GS) or random search [8][9]. Some manual methods are also proposed, for example: the number of hidden neurons is 2/3 of the number of inputs plus the number of outputs, the number of hidden neurons per hidden layer follows the rule of pyramid geometry, the relationship between the number of hidden layers and hidden layers is logarithmic [10]. In some the literatures mentioned above, there is no clear guideline on how to determine the number of hidden layers or neurons optimally, even the roles and functions of both are explained minimally. Some literatures propose methods or ways to determine the number of hidden layers or neurons, but they are not generally applied. It depends on the type of the input and output data. Researches in this area still leaves difficult research tasks [11]. Some of the methods mentioned apparently cannot be applied to different types of data. Some researchers determined the number of hidden layers or neurons based on their past experience, while beginner researchers even did it with 'trial and error' method. Although it is still a difficult area of research, researches to determine the number of hidden layers and neurons must continue to be carried out, because

978-1-5386-8448-1/19/$31.00 ©2019 IEEE

these two will greatly determine the deep neural network learning performance.

The development of methods for determining the number of hidden layers and neurons in the neural network can be explained from a philosophical perspective by using several key concepts of Kuhn's thoughts related to his theory of the structure of scientific revolution in the book entitled The Structure of Scientific Revolution published in 1962. This theory is a new offer for scientific discourse based on the history of the development of science [12][14].

As it does not yet have established guidelines in determining the neural network topology, and each of the method developed is only applied to certain datasets, it is difficult to do a verification test with the ultimate goal of generalizing to existing methods, as well as to new methods that will be developed later. One test that can be done to find out whether the methods developed to determine the neural network topology is true or not is to use the theory of falsification that was put forward by Karl Raymond Popper [13][14]. Popper's theory asserts that the truth of the proposition of a science is not determined through verification tests, but through an attempt to deny the truth through various systematic experiments. The greater the effort to deny a theory, and if the theory turns out to continue to be able to survive, the more solid its existence will be..

II. LITERATURE REVIEW

A. *Thomas Kuhn's Philosophy*

Kuhn offerred a new theory related to science that refers to the process of the scientific development rather than the product it produces. Kuhn tried to shift the subject of science to the activity of science (to produce). Thus, he shifted logical analysis and explanation of a science as a product that has been established to a natural or historical explanation of the scientific process[12].

Paradigm

One of the key words when speaking of Thomas Kuhn's thinking is "paradigm". This idea would like to emphasize that scientific theory is not only limited to a set of theoretical principles, but also includes world views in science, which is what Kuhn then initiated as a "paradigm".

The Science Revolution

According to Kuhn, science developed revolutionary from one paradigm to another. The Structure of Scientific Revolution book contains the stages of scientific revolution meant by Kuhn. Thomas Kuhn mapped the stages of the development of science into four main phases.

Pre-paradigm Phase

This phase is also called as the immature science phase. This phase is a period that takes a long time. Here, scientific researches on certain things are carried out without specific directions and purposes. This period also emerges various kinds of thoughts that compete with each other and exclude each other. It has different conceptions about the basic problems of scientific discipline and what criteria should be used to evaluate theories.

Normal Science Phase

To become a science, a scientific discipline must reach a consensus that is in the shade of a particular paradigm. Of the various sciences that developed in the pre-paradigm phase, one thought or theory will emerge which then dominates other theoretical or scientific disciplines. Schools or other thoughts are oriented and recognize the superiority of the dominant school or thought. In this case, it promises more accurate problem solving and a more advanced future research so that it is more dominant than its competitors.

When a consensus has been reached, Kuhn claimed that scientists had begun to get into normal sciences. The normal science precondition is that there is a commitment to the existence of a shared paradigm that will determine the rules of the game and all standard benchmarks in scientific practice. "Normal" scientists will not make new discoveries outside the prevailing paradigm. Instead, they are fully involved in using the paradigm to better understand the symptoms of natural symptoms in more detail.

Anomaly and Crisis Phase

This phase is also called as the phase of the emergence of extraordinary sciences. At this time, knowledge, both in examples of scientific practice (copies) and disciplinary matrices, can no longer be relied upon in solving problems that arise. The emergence of a very crucial and unsolved problem does not only make scientists confused, but it also creates a crisis in the scientific community. Since then, they began to question the prevailing paradigm.

The Emergence of a New Paradigm Phase

In the midst of competition during a crisis, one of the emerging thoughts will be able to overcome scientific problems and then be able to generalize and promise the future of better scientific researches. At this point, extraordinary sciences become normal sciences. This change is the climax of Kuhn's scientific revolution. He explained this as "an episode of non-cumulative development in which an older paradigm is replaced in whole or in part by a new paradigm that is more compatible."

B. *Popper's Philosophy*

According to Popper, a theory or proposition of science or knowledge is not seen as scientific only because it can be verified by verification as the scientists think, but because it can be tested (testable) through various systematic experiments to deny it (falsification). If a hypothesis or a theory can survive against all denials, then the truth of the hypothesis or theory is further strengthened. He calls it as corroboration. The greater the effort to deny a theory, and if the theory turns out to continue to be able to survive, the more solid its existence will be.

Furthermore, Popper explained that every scientific theory is always hypothetical, in the form of conjecture, there will never be a final truth. Every theory is always open to be replaced by a new theory that is more appropriate. Related to this, he preferred to use it with the term hypothesis rather than theory, only solely based on the nature of its temporality. He asserted that a hypothesis or proposition is said to be scientific if in principle it has the possibility to be denied (refutability) [13].

III. DISCUSSION

This section describes the development of methods for determining the neural network topology by using Thomas Kuhn's philosophy and on how to test the correctness of methods by using Popper's theory of falsification..

A. The Development of the Neural Network Topology Determination Method from Kuhn's Point of View

Pre-Paradigm Phase

From the studies determining the neural network topology that have been done, the phase can be categorized into three:

1. Researches that focus only on determining the number of hidden neurons in one hidden layer [15][16][17]
2. Researches that focus on comparing the use of one hidden layer and two hidden layers [18]
3. Researches that focus on determining both number of hidden layers and hidden neurons [19][1]

The first group of researches is carried out with the assumption that the use of one hidden layer is able to approach almost all functions. In addition, the majority of studies in this group does not consider the characteristics of the input feature, but only pay attention to the number of features or amount of data. The second group of researches is conducted by looking at the opportunity that the use of two hidden layers could improve the network performance in line with the increase in computer capabilities. The comparison results are obtained without looking at the characteristics of the input features, even though the characteristics of the input features can be considered when deciding whether to only use a hidden layer or multilayer. The third group of researches focuses on determining the number of hidden layers and neurons at once. The majority of the researches are conducted by trial and error method or using rule of thumb for experienced researchers. Therefore, this group does not pay attention to the characteristics of the input feature to determine the network topology. The input feature character that can be considered is by calculating the correlation or variation between or intra input features.

Of the three research groups, almost all the proposed methods cannot be used as a guideline for determining the right neural network topology for other researchers. Consequently, based on Kuhn's thinking, the method for determining neural network topology is still in the pre-paradigm phase.

Normal Science Phase

Although in the pre-paradigm phase, there are some researchers who try to determine the neural network topology that can be more general. For example, as was done by Tej and Holban [20], who tried to determine the neural network topology by using clustering and regression method. The method developed has begun to consider the characteristics of the dataset's input features. However, because it is still a new research and the results have not been tested by other researchers, it still needs time to test whether the method proposed by Tej and Holban can be normal science..

B. Testing the Determination of the Neural Network Topology Method from the Perspective of Popper

Several studies conducted before Tej and Holban have been difficult to apply by other researchers for different datasets so that they are difficult to be falsified, whereas the research conducted by Tej and Holban must be systematically tested to be denied. The tests to deny the method proposed by Tej and Holban can be done with the following scenario:

- Testing the method proposed by Tej and Holban by using certain datasets that have objective functions for predictions
- Testing the method proposed by Tej and Holban by using certain datasets that have objective functions for classification
- Comparing the neural network topology obtained by the method proposed by Tej and Holban with several other neural network topologies, whether the topology proposed by Tej and Holban provides better performance or not.

If the proposed method can stand up against all denials, then the truth of the method is increasingly strengthened, or what Popper calls as corroboration.

IV. CONCLUSION

From the explanations in the previous chapters, it can be concluded that there are three researches groups in determining the neural network topology, namely: researches that only focus on the number of hidden neurons in one hidden layer, researches that focus on the comparison of one hidden layer and two hidden layers, and researches that focus on determining both number of hidden layers and neurons. None of the three researches groups can be used as a guide in determining the neural network topology for other researchers. Therefore, according to Kuhn's thought, researches in the area of determining neural network topology are still in the pre-paradigm phase. New researches conducted in this area are being directed so that it can be generally accepted as normal sciences. Popper's theory of falsification can be used to test new methods produced.

REFERENCES

[1] Qolomany, B., Maabreh, M., Al-Fuqaha, A., Gupta, A., & Benhaddou, D. (2017). Parameters optimization of deep learning models using Particle swarm optimization. 2017 13th International Wireless Communications and Mobile Computing Conference, IWCMC 2017, (Iwcmc), 1285–1290. https://doi.org/10.1109/IWCMC.2017.7986470

[2] Nitta, T. (2017). Resolution of Singularities Introduced by Hierarchical Structure in Deep Neural Networks. IEEE Transactions on Neural Networks and Learning Systems, 28(10), 2282–2293. https://doi.org/10.1109/TNNLS.2016.2580741

[3] Madhiarasan, M. & Deepa, S. N. (2018). A Novel Method to Select Hidden Neurons in ELMAN Neural Network for Wind Speed Prediction Application, WSEAS Transactions On Power Systems, 13, 13–30.

[4] Vujičić, T., & Matijevi, T. (2016). Comparative Analysis of Methods for Determining Number of Hidden Neurons in Artificial Neural Network. Central European Conference on Information and Intelligent Systems, 219.

[5] Thomas, L., Manoj Kumar, M. V., & Annappa, B. (2016). Discovery of optimal neurons and hidden layers in feed-forward Neural Network. 2016 IEEE International Conference on Emerging

Technologies and Innovative Business Practices for the Transformation of Societies, EmergiTech 2016, 286–291. https://doi.org/10.1109/EmergiTech.2016.7737353

[6] Thomas, A. J., Petridis, M., Walters, S. D., Gheytassi, S. M., & Morgan, R. E. (2016). On predicting the optimal number of hidden nodes. Proceedings - 2015 International Conference on Computational Science and Computational Intelligence, CSCI 2015, 565–570. https://doi.org/10.1109/CSCI.2015.33

[7] Lorenzo, P. R., Nalepa, J., Kawulok, M., Ramos, L. S., & Ranilla, J. (2017). Particle Swarm Optimization for Hyper-Parameter Selection in Deep Neural Networks. In Proceedings of GECCO '17, Berlin, Germany, July 15-19, 2017, 8 pages

[8] Bergstra, J., & Yoshua Bengio, U. (2012). Random Search for Hyper-Parameter Optimization. Journal of Machine Learning Research, 13, 281–305. https://doi.org/10.1162/153244303322533223

[9] Larochelle, H., Courville, A., & Bergstra, J. (2007). An Empirical Evaluation of Deep Architectures on Problems with Many Factors of Variation, Proceedings of the 24th International Confer- ence on Machine Learning.

[10] Grabusts, P., & Zorins, A. (2015). The Influence of Hidden Neurons Factor on Neural Network Training Quality Assurance, 3, 76–81.

[11] Lee, S., Ha, J., Zokhirova, M., Moon, H., & Lee, J. (2018). Background Information of Deep Learning for Structural Engineering. Archives of Computational Methods in Engineering, 25(1), 121–129. https://doi.org/10.1007/s11831-017-9237-0

[12] Kuhn, T. S. (2002). The Structure of Scientific Revolution. Translate: Tjun Surjaman. Bandung: Rosda Karya.

[13] Popper, Karl R. (2005). The Logic of Scientific Discovery. New York: Rouledge.

[14] Mahayana, Dimitri (2018). Philosophy of Science. Lecture Note. Bandung: Penerbit ITB.

[15] Hunter, D., Yu, H., Pukish, M. S., Kolbusz, J., & Wilamowski, B. M. (2012). Selection of Proper Neural Network Sizes and Architectures — A Comparative Study. IEEE Transactions on Industrial Informatics, 8(2), 228–240. https://doi.org/10.1109/TII.2012.2187914

[16] Madhiarasan, M., & Deepa, S. N. (2016). A novel criterion to select hidden neuron numbers in improved back propagation networks for wind speed forecasting. Applied Intelligence, 44(4), 878–893. https://doi.org/10.1007/s10489-015-0737-z

[17] Madhiarasan, M., & Deepa, S. N. (2017). Comparative analysis on hidden neurons estimation in multi layer perceptron neural networks for wind speed forecasting. Artificial Intelligence Review, 48(4), 449–471. https://doi.org/10.1007/s10462-016-9506-6

[18] Guliyev, N. J., & Ismailov, V. E. (2018). On the approximation by single hidden layer feedforward neural networks with fixed weights. Neural Networks, 98, 296–304. https://doi.org/10.1016/j.neunet.2017.12.007

[19] Koutsoukas, A., Monaghan, K. J., Li, X., & Huan, J. (2017). Deep-learning: Investigating deep neural networks hyper-parameters and comparison of performance to shallow methods for modeling bioactivity data. Journal of Cheminformatics, 9(1), 1–13. https://doi.org/10.1186/s13321-017-0226-y

[20] Tej, M. L., & Holban, S. (2018). Determining Optimal Neural Network Architecture Using Regression Methods. 2018 International Conference on Development and Application Systems (DAS), 180–189.