Scholar Publishing Group
International Journal of Neural Network
https://doi.org/10.38007/NN.2020.010106
ISSN 2790-2692 Vol. 1, Issue 1: 41-48

Neural Network Stability Fusing Robust Features

Edris Asghari*

Gachon University, Republic of Korea

**corresponding author*

Keywords: Robust Feature, Neural Network, Stability Study, Network Performance

Abstract: The current successful application of deep learning is based on Deep Neural Network (DNN). Robustness can help users obtain the service quality information of neural network (NN) in practical applications, measure the security of NN, and avoid potential security threats. In the existing robust computing research, there is no method that can give the robustness of a NN for a certain input sample in a timely and effective manner in practical applications. Therefore, in this paper, the robustness features are combined to study the stability of NN. This paper firstly describes the stability and robustness evaluation framework of NN, and then studies the stability of NN from three aspects: robust classification, bifurcation threshold and robustness predictor stability. Robustness indicators and network performance are analyzed and corresponding conclusions are drawn.

1. Introduction

The rapid development of NN technology plays a significant role in promoting the development of artificial intelligence, and also promotes the industrialization of deep learning [1-2]. NN are vulnerable to adversarial samples, but there is still no effective defense method to protect NN from adversarial samples, so the research on NN security and quality assurance is extremely important [3]. Robustness is an important attribute to measure the security and reliability of NN models, which is of great significance in the research of NN security assurance [4-5]. In DNN, the stability of forward propagation is related to the representation stability and generalization ability of the net, especially the adversarial attack problem of DNN [6].

In recent years, many scholars have conducted in-depth research on robustness in order to solve the problem of NN stability, and have achieved good results. For example, scholars such as Strisciuglio N use the kernel object called graphons to analyze the stability of GNN, and study the stability of graphon disturbance based on graphon signal processing theory. It can be seen from observation that GNN is stable to graphon disturbance, and its stability The sexual bounds decrease asymptotically with the size of the graph [7]. Researchers such as Szilassy P proposed a new

neighborhood preservation layer to improve the robustness of the network by replacing the fully connected layers. The NN architecture contains these layers and can be trained efficiently. It is theoretically proved that the NN architecture is suitable for State-of-the-art gradient descent based attacks are more robust [8]. Robustness is very important for the stability study of NN.

The challenges faced by the current deep learning include the robustness challenge brought by the data samples, so this paper studies the robustness of the NN by combining the robustness features [9-10]. The structure of this paper can be roughly divided into three parts: the first part is an overview of NN related theories, including the stability of NN and the robustness evaluation framework; the second part is stability research, mainly robust The three aspects of the classification, the bifurcation threshold and the stability of the robust predictor are studied. It is found that the purpose of bifurcation control can be achieved by adjusting the connection weight of the NN. The third part is the analysis of research results, including robustness index analysis and network performance analysis, attractor dynamics can enhance network stability and improve network generalization ability.

2. Related Overview

2.1. Stability of NN

In the feature extraction process of DNN (DNN), we regard the input as an initial state of the feature space, and our main concern is how to transform the input state into features in a subset of the feature space, so that the target task can make People satisfactorily utilize the learned feature representation [11]. This subset can be an invariant set to ensure stable feature representation [12]. Here, we further assume that these sets are attractors, using attractors to enhance the stability of NN forward propagation.

The good forward stability of the network means that the features extracted by the network have appropriate changes to the changes of the network input, and will not have drastic changes due to slight changes in the input of the DNN, resulting in unstable predictions of the network [13-14]. For example, in an image, a slight change in the content of the image does not affect the meaning it expresses, the slight movement of the object in the image, the slight change in the light and shadow in the image, the slight change in the pose of the object, etc., should not make the features extracted by the DNN occur. A larger change than the features extracted before the change [15].

2.2. Robustness Evaluation Framework

Conflict robustness is a measure of the model's robustness to adversarial samples, and in the real environment, there are also some reasonable abnormal data, such as sample data generated by natural transformations such as changing brightness, contrast, rotation, translation, blur, and zoom. [16]. Compared with the adversarial samples obtained by artificially adding malicious perturbations, these sample data are more likely to appear in the real environment of model deployment, and have important practical significance for the robustness evaluation of the model. In this paper, these samples are called natural transformation samples. A robust NN not only needs to perform well on the original samples, but also needs to ensure high performance on some naturally transformed samples [17-18]. Combined with the above analysis, this paper designs an evaluation framework as shown in Figure 1. The framework mainly includes three modules: test data generation, robustness evaluation and evaluation conclusion.

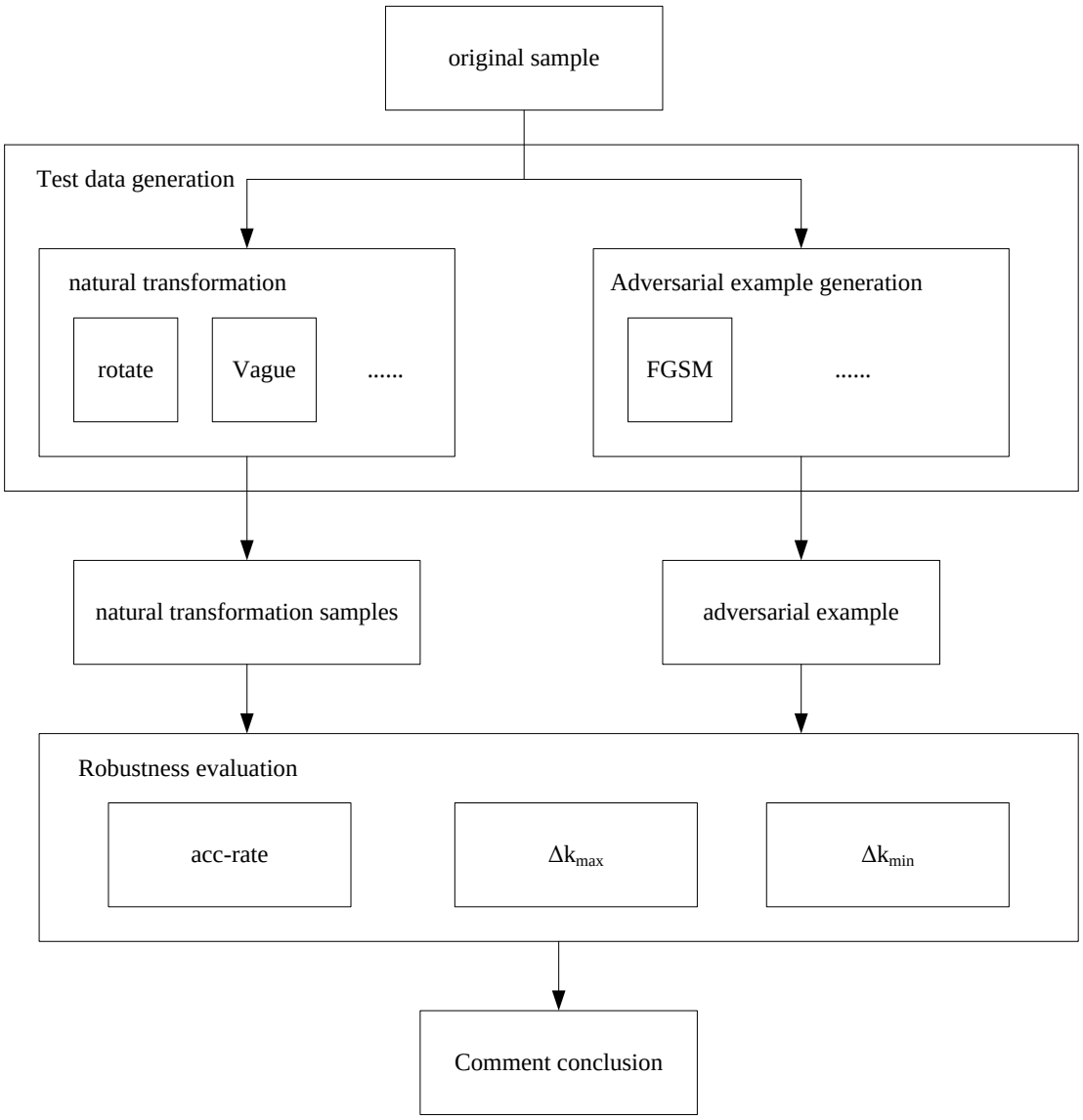

Figure 1. Robustness evaluation framework

3. Stability Study

3.1. Robust Classification

The robustness of NN can be roughly divided into two types, namely global robustness and local robustness.

The global robustness requires that under the distance constraint of δ, for the distance between two points b1 and b2 in the input space to be less than L2, then the output gap of these two samples in the NN model N will be less than the given θ. As shown in the formula below, where $\forall b_1, b_2 \in D$:

$$\|b_1 - b_2\| \le \delta \Rightarrow \forall \lambda \in L \tag{1}$$

$$\|N(b_1, \lambda) - N(b_2, \lambda)\| < \theta \tag{2}$$

Since the constraints of global robustness are too strong, we generally relax the constraints to consider local robustness. Local robustness requires that given the initial sample point b0, under the distance constraint δ, the points satisfying $\|b-b_0\| \le \delta$ in the input space will be classified into the same category as b0. The formula is as follows:

$$\forall b, \|b-b_0\| \le \delta \Rightarrow L(b,N) = L(b_0,N) \tag{3}$$

In this paper, the adversarial examples we consider violate local robustness. Because the given initial sample point b0 is under a certain distance constraint δ, we can add perturbation to b0 to obtain b0' such that $\left\|b_0' - b_0\right\| \le \delta$. But b0' and b0 are classified into different categories. Therefore, the robustness we hope to improve in this paper belongs to a kind of local robustness.

3.2. Bifurcation Threshold

This section investigates the effect of changes in bifurcation threshold weights on the stability of NN. First, the effect of weight changes on the bifurcation threshold is selected with suppressive properties. It can be seen from the analysis in Figure 2 that when other parameters are fixed and d1=d2=d, this section finds that the bifurcation threshold d0 will gradually increase with the change of v2, but the bifurcation threshold will not always exist. When the value v2 is -1.6, the NN bifurcation threshold will always remain asymptotically stable at the origin. Similarly, increasing the connection weight v4 also expands the bifurcation threshold of the NN. However, the smaller the absolute value of the two weights, the smaller the bifurcation threshold of the NN and the smaller the stable interval of the NN. Therefore, the purpose of bifurcation control can be achieved by adjusting the connection weight of the NN.

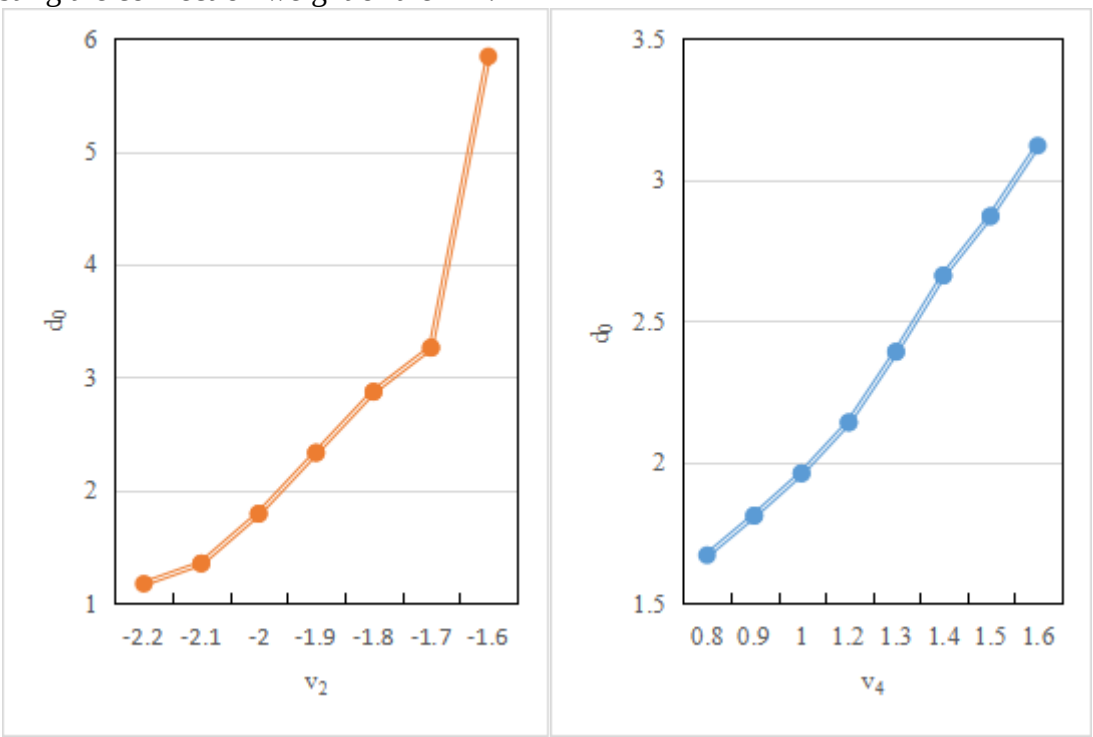

Figure 2. Effect of weight change on bifurcation threshold

3.3. Stability of Robust Predictors

Generally speaking, the training process of NN requires a lot of skills and experience, and users cannot guarantee that the model achieves the best performance during the process of applying the NN, and the trained model may have different accuracy rates. For models with different accuracy rates, their robustness performance will also be different. The accuracy rates of the LeNet-5 model on the MNIST dataset are 78.27%, 83.56%, 88.91%, 93.69%, and 97.85%, respectively, and the robustness mean of the test samples is calculated, and the robustness mean calculation results are: 1.88, 1.57, 1.46, 1.42, 1.37. As can be seen from Figure 3, as the accuracy of the model increases, the average robustness shows a downward trend, but the decline gradually becomes flat. The results show that the robustness predictor can effectively predict the magnitude of robustness for NN models with different accuracy rates.

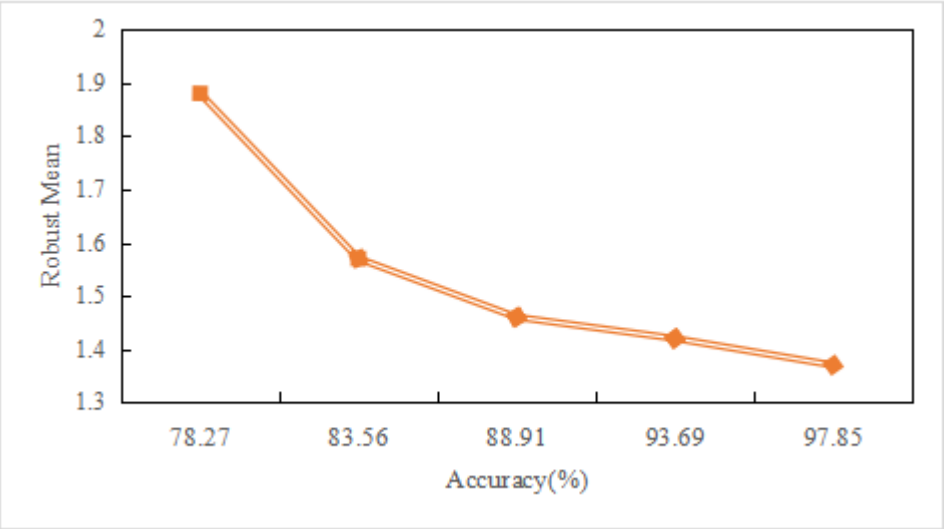

Figure 3. NN model robustness prediction accuracy

4. Research Analysis

4.1. Analysis of Robustness Index

In this experiment, the MNIST and Fashion-MNIST datasets are used. The trained models are lenet-1, lenet-4 and 1enet-5. The test data is still generated using the FGSM method. For the original MNIST test set, the upper limit of the perturbation intensity is set as 0.06; for the Fashion-MNIST original test set, set the perturbation upper limit to 0.03, and generate 5 adversarial samples for each original test sample of each dataset as test data for robustness evaluation. Input the test data into the corresponding model to calculate the conflict robustness index Δkmin, and to verify the validity of the index, calculate the L2 distance between each adversarial sample and its original sample according to formula (4), and take the minimum value, and finally Find the average.

$$r(a,a') = \sqrt{\sum_{i=1}^{N}(a_i' - a_i)^2} \tag{4}$$

where a represents the original sample, and a' represents the adversarial sample. The calculated Δkmin and $\overline{L_2}$ are counted, and the calculation results of the indicators on the MNIST dataset are shown in Table 1.

Table 1. Indicator calculation results

Data set	Model	Δk_{min}	$\bar{L}_2$
MNIST	lenet-1	0.1478	2.8147
	lenet-4	0.4567	2.8378
	lenet-5	0.6791	2.8653
Fashion-MNIST	lenet-1	0.0768	0.6789
	lenet-4	0.1498	0.7654
	lenet-5	0.2014	1.4925

From Table 1, it can be observed that under the FGSM attack, among the models trained on the MNIST dataset, the Δkmin of the lenet-5 model is 0.6791, the lenet-4 model is 0.4567, and the minimum lenet-1 model is 0.1478. The average minimum distance L2 of the adversarial sample from the original sample is also the same size relationship. Because under the same attack algorithm, the larger the required disturbance, the better the defense effect of the model, that is, the stronger the model robustness, so the robustness of the lenet-5 model is stronger than that of the lenet-4 model, and the lenet-4 model The robustness is stronger than the lenet-1 model. This is consistent with the results of the model robustness evaluation using the conflict robustness metric Δkmin. Among the models trained on the Fashion-MNIST dataset, the Δkmin for the lenet-5 model is 0.2014, which is larger than 0.1498 for the lenet-4 model and 0.0768 for the lenet-1 model. The average minimum distance L2 between the original sample and the adversarial sample is also the same size relationship. Therefore, the same robustness evaluation conclusion can be drawn using the conflict robustness index Δkmin and the average minimum distance L2, the robustness of the lenet-5 model is stronger than that of the lenet-4 model, and the robustness of the lenet-4 model is stronger than that of the lenet-4 model. lenet-1 model. Through the analysis of the above two sets of experimental results, it can be proved that it is effective to use the conflict robustness index Δkmin to evaluate the robustness of the DNN model. The larger the Δkmin, the better the model robustness.

4.2. Network Performance Analysis

Table 2 lists the results of testing with different numbers of initial placement attractors on different datasets and web structures. Each set of experiments was tested 5 times. The table shows the mean and standard deviation of the tests. In Table 2, h=0 indicates that the net training is not connected to RMAN. From the analysis in Table 2, it can be seen that when RMAN is used, the gap between exercise loss and test loss is smaller than that without RMAN (i.e., h=0), resulting in better generalization performance. This suggests that enhancing web stability based on attractor dynamics is beneficial for web performance and can improve the generalization ability of the net.

Table 2. NN test error rate comparison

Dataset	Model	Layer	Number of placement attractors			
			h=0	16	64	256
CIFAR10	ResNet	22	8.54	8.43	8.01	7.98
	ResNet	35	7.45	6.75	7.13	6.87
CIFAR100	ResNet	39	33.64	31.96	32.91	31.58
	ResNet	62	32.78	32.04	31.23	30.45
	LM-ResNet	123	30.97	30.09	29.97	29.53
	LM-ResNet	184	28.14	27.31	27.34	26.36

5. Conclusion

The robustness of the NN is an important index to evaluate the security of the NN, and it is of

great significance in ensuring the security of the NN. Therefore, this paper studies the stability of the NN based on the robustness characteristics. In view of the current research status of NN robustness, this paper proposes a robustness predictor, which can timely and effectively give the robustness information for a specific NN to be tested, and provide quality and safety assurance for the NN. Serve. In this paper, it is found through research that enhancing network stability based on attractor dynamics is beneficial to net performance and can improve the generalization ability of the web. There are still many shortcomings in this paper that need to be improved.

Funding

This article is not supported by any foundation.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

[1] Bondarev N V. Exploratory, Regression, and Neural Network Analysis of the Stability of Cation Coronates in Selected Pure Solvents. Russian Journal of General Chemistry, 2020, 90(10):1906-1920. https://doi.org/10.1134/S107036322010014X

[2] Woods W, Chen J, Teuscher C. Adversarial explanations for understanding image classification decisions and improved neural network robustness. Nature Machine Intelligence, 2019, 1(11):508-516. https://doi.org/10.1038/s42256-019-0104-6

[3] Kiss A, Bischof B. Robustness Testing of Neural Network for Handwritten Digit Recognition. Alkalmazott Matematikai Lapok, 2020, 37(1):87-101. https://doi.org/10.37070/AML.2020.37.1.04

[4] Gazizov M R, Grigorian K A. Of Neural Network Model Robustness Through Generating Invariant to Attributes Embeddings. Russian Digital Libraries Journal, 2020, 23(6):1142-1154. https://doi.org/10.26907/1562-5419-2020-23-6-1142-1154

[5] Ilina O V, Tereshonok M V. Robustness Study of a Deep Convolutional Neural Network for Vehicle Detection in Aerial Imagery. Journal of Communications Technology and Electronics, 2020, 67(2):164-170. https://doi.org/10.1134/S1064226922020048

[6] Nguyen D A, Minh K D, Nguyen M L, et al. A symbolic execution-based method to perform untargeted attack on feed-forward neural networks. Automated Software Engineering, 2020, 29(2):1-29. https://doi.org/10.1007/s10515-022-00345-x

[7] Strisciuglio N, Lopez-Antequera M, Petkov N. Enhanced robustness of convolutional networks with a push–pull inhibition layer. Neural Computing and Applications, 2020, 32(3):1-15. https://doi.org/10.1007/s00521-020-04751-8

[8] Szilassy P, B Néneth, P Gáspár. Design and robustness analysis of autonomous vehicles in intersections. IFAC-PapersOnLine, 2019, 52(8):321-326. https://doi.org/10.1016/j.ifacol.2019.08.090

[9] Kumar R, Srivastava S. Comparative study of neural networks for dynamic nonlinear systems identification. Soft Computing, 2019, 23(4):1-14. https://doi.org/10.1007/s00500-018-3235-5

[10] Yashas S, Raghunathan S, Priyakumar U D. SCONES: Self-Consistent Neural Network for Protein Stability Prediction Upon Mutation. The Journal of Physical Chemistry B, 2020, 125(38):10657-10671. https://doi.org/10.1021/acs.jpcb.1c04913

[11] Sene N. Fractional input stability and its application to neural network. Discrete and Continuous Dynamical Systems - Series S, 2019, 13(3):853-865. https://doi.org/10.3934/dcdss.2020049

[12] Aljarbouh A, Sharam A, Khateeb M. Using Time-Delay Neural Network (DTDNN) for Enhancement Power Systems Dynamic Stability. International Journal of Computer Applications, 2020, 176(12):40-47. https://doi.org/10.5120/ijca2020920080

[13] Gobinath S, Madheswaran M. Deep perceptron neural network with fuzzy PID controller for speed control and stability analysis of BLDC motor. Soft Computing, 2020, 24(13):10161-10180. https://doi.org/10.1007/s00500-019-04532-z

[14] Alade O, Mahmoud M, Shehri D A, et al. Rapid Determination of Emulsion Stability Using Turbidity Measurement Incorporating Artificial Neural Network (ANN): Experimental Validation Using Video/Optical Microscopy and Kinetic Modeling.. ACS omega, 2020, 6(8):5910-5920. https://doi.org/10.1021/acsomega.1c00017

[15] Nguyen H H, Zieger T, Braatz R D, et al. Robust Control Theory Based Stability Certificates for Neural Network Approximated Nonlinear Model Predictive Control. IFAC-PapersOnLine, 2020, 54(6):347-352. https://doi.org/10.1016/j.ifacol.2020.08.568

[16] Martynyuk A, Stamova I, Martynyuk-Chernienko Y A. On the boundedness and Lagrange stability of fractional-like neural network-based quasilinear systems. The European Physical Journal Special Topics, 2020, 231(10):1789-1799. https://doi.org/10.1140/epjs/s11734-022-00447-3

[17] Nguyen H, Shang L, Bui X N, et al. Toward state-of-the-art techniques in predicting and controlling slope stability in open-pit mines based on limit equilibrium analysis, radial basis function neural network, and brainstorm optimization. Acta Geotechnica, 2020, 17(4):1295-1314. https://doi.org/10.1007/s11440-021-01373-9

[18] Jawa B, Li F, Xywa B, et al. Enhanced stability results for generalized neural networks with time-varying delay - ScienceDirect. Journal of the Franklin Institute, 2020, 357(11):6932-6950. https://doi.org/10.1016/j.jfranklin.2020.04.049