Influence and optimization of neural network topology on steel performance prediction model

Pengtao Wang*

*Research Institute of Technology, Shougang group co. LTD, Beijing, 100041,China
*Corresponding author’s e-mail: wangpengtao@shougang.com.cn

Abstract. Deep neural networks have powerful nonlinear performance, but until now there is no ideal method for hyperparameters tuning. In this paper, prediction correct rate of IF steel mechanical performance is used as evaluation criterion to establish a neural network regression model. The study shows that increasing the number of hidden layers and neurons could both improve the prediction accuracy. Under the same parameter magnitude, the effect of adding hidden layer numbers is better than neuron numbers. However, after the network depth reaches a certain threshold, the accuracy does not increase, or even drop. After a lot of attempts, we have obtained a model topology of 3 hidden layers and 100 neurons with 1 output feather. On the new test set of 1,000 samples, the yield strength(Rp0.2) tensile strength(Rm) and elongation(A80) prediction accuracy are 74.4%, 89.8% and 83.2%. Data expansion, test method stabilization and adding knowledge-driven sub-model could be used to optimize the model.

1. Introduction

Steel production is continuous and multi-process. Any small changes in raw materials, smelting, hot rolling, cold rolling, annealing, coating and other processes, could have great impact on the final product, which are difficult to control. In modern steel enterprises, to deal with order demand of different performance and product specifications, experienced technical engineers usually need to spend much time to evaluate, analyze, and develop different processes. After that, the final performance needs to be tested before delivering to customers. All these work consume a lot of manpower and time resources, but still do not solve the product performance fluctuations effectively, especially for high quality materials such as automotive steel. The unstable performance will lead to a lower production efficiency and high scrap rate[1, 2].

Regression neural network is suitable for solving this kind of complex and non-linear issue. By importing a large amount of production data, appropriate algorithms and evaluation criterion, it helps to provide real-time performance reference, which is beneficial to optimize and improve product quality[3-5].

Topology plays an important role on neural network. With improvements of computing power of CPU/GPU and optimized algorithms, deep networks show strong ability to solve complicated issues on regression and classification. In 2012, the 8-layer deep neural network was released, achieving huge performance improvements in the image recognition competition[6]. Since then, dozens and hundreds of layers of deep neural networks have been proposed and applied. Although some research theories have proved that if appropriate neuron numbers in the hidden layer is selected, a three-layer neural network plus the activation function can infinitely approximate any continuous function with
any precision[7, 8]. But it is not an easy task to choose the right number of neurons in a three-layer neural network.

In this paper, through Python language and Tensorflow deep learning framework, we construct a regression neural network, focusing on three hyperparameters factors: hidden layer neurons, hidden layers and output layer neurons and their impacts on performance prediction accuracy in new data set. Subsequently, the paper has a discussion about model optimization.

2. Neural network construction

2.1. Product selection

China has an annual automotive production and sales volume for more than 20 million, and the average amount of steel used for each car exceeds 500 kg. According to scope of use, it can be divided into two categories: mild steel used in covering parts, and high strength steel in structural parts. Compared with the varieties of high strength steel which consists of DP, TRIP, HSLA, HS, etc., IF steel is widely used in stamping parts due to its excellent formability and welding performance. And for the body-in-white of most passenger cars, it is also the most used material. This helps us to obtain a large amount of labeled data, which is one of the key factors for supervised learning. Therefore, we choose IF steel as the data source.

2.2. Data set

The process and product data set in this paper are provided by Shougang group, which consists of 6,000 samples. Considering that inappropriate data set division may result in different distributions in train and validation set, which increases the risk of overfitting. That is to say, the prediction accuracy in validation set is much lower than train set[9]. Here we use K fold cross validation to divide the data set into 6 equal parts, and each of them contains 1000 samples[10-12]. Four of them are train sets, the other one is validation set. This process loops 5 times. When the training is over, a test set of 1000 new data are used for model evaluation.

2.3. Input feathers

Among the process parameters that affect mechanical properties of IF steel, chemical composition is the most important one. IF steel substrate has no interstitial atoms such as C and N, but that does not mean these elements are nonexistent, which are fixed by strong carbide or nitride elements such as Nb, Ti, and Al, forming Ti(C,N)\(_x\), Nb(C,N)\(_x\), Ti(C,S)\(_x\) or AlN. These precipitates will aggregate and grow in hot rolling, coiling, and annealing processes, with different sizes and distributions. During the subsequent deformation, they interact with material matrix, having different effects on the properties. Mn, Si, and P elements are strengthening elements, especially for P. Rolling and coiling also have great impacts on properties[13, 14]. We generally use a high cooling temperature to make the precipitate growing as larger as possible during cooling, which could reduce the influence of fine grain and precipitation on the performance enhancement, and obtain high plasticity. The annealing
temperature is also essential to ensure that grains are recovered and recrystallized. In general, the flat elongation only affects the yield strength and elongation. Unless a certain threshold is reached, the tensile strength will be significantly affected. Based on the analysis above, all 21 input feathers are listed in Table 1.

Table 1. Input feathers of chemical elements and processes.

| Category                  | Content                                                                 |
|---------------------------|------------------------------------------------------------------------|
| Chemical element          | C, Si, Mn, P, S, Al, Nb, Ti, N, B                                       |
| Hot rolling               | heating temperature, initial rolling temperature, final rolling temperature, reduction rate, coiling temperature |
| Cold rolling and annealing| reduction rate, annealing temperature, slow cooling temperature, rapid cooling temperature, over aging temperature, flat elongation |

2.4. Data standardization
In order to ensure better convergence effect, we use equation (1) for data normalization, which eliminates dimensions and unifies the extreme value of input feathers ranging from 0 to 1 [9].

\[
x_i = \frac{x - \min(x)}{\max(x) - \min(x)}
\]

(1)

2.5. Output feathers
IF is a kind of steel for deep drawing, that plastic strain ratio (r) and strain hardening exponent (n) are two important parameters to evaluate the stamping performance. It seems that we need to add r and n to the conventional three mechanical properties. However, r and n is highly correlated with the plastic elongation of the material, especially for n. It can be considered that the plastic elongation is high, so is n as well [15]. And the measurement method (such as selecting different ranges) has a great influence on the r and n. In addition, adding an output feature will increase the parameters of the entire network. Therefore, we only focus on yield strength (Rp0.2), tensile strength (Rm) and elongation (A80).

2.6. Evaluation criteria
Unlike classification issue, product performance is not a discrete variable of 0 or 1, which is a continuous variable. And due to test error, even for the same sample, the test results may not be the same. Therefore, it is necessary to comprehensively consider the size of the measurement error and the difference in performance to formulate the evaluation criteria thresholds.

Table 2. Evaluation criteria for Results.

| Output feathers | OK                | NG                               |
|-----------------|-------------------|----------------------------------|
| Rp0.2(MPa)      | [-3,+3]           | (−∞,−3) U (+3,+∞)               |
| Rm(MPa)         | [-5,+5]           | (−∞,−5) U (+5,+∞)               |
| A80(%)          | [-2,+2]           | (−∞,−2) U (+2,+∞)               |

2.7. Activation function
Adopting the Relu activation function could not only guarantee the non-linear unit, but also reduce the computational complexity of each parameter updating [16, 17].

2.8. Algorithm
For multi-dimensional regression issue, BP network with nonlinear activation function is the most effective combination. We use mini batch stochastic gradient descent algorithm and MSE loss function. And at the t time iteration, the matrix of parameter \( W \) and bias \( b \) are updated by equation (2) and (3) [18].
\[ W_i = W_{i-1} - \eta \sum_{t} \frac{\partial E_{t-1}}{\partial W_{i-1}} \]
\[ b_i = b_{i-1} - \eta \sum_{t} \frac{\partial E_{t-1}}{\partial b_{i-1}} \]  

2.9. Optimizer

In order to ensure a higher convergence rate, the model realizes dynamic adjustment of learning rate through Adam optimizer. And the attenuation rates \( \beta_1 \) and \( \beta_2 \) usually are 0.9 and 0.99\[19, 20\].

\[ M_i = \beta_1 M_{i-1} + (1-\beta_1)g_i \]
\[ G_i = \beta_2 G_{i-1} + (1-\beta_2)g_i \]  

2.10. Regularization

For the purpose of promoting the generalization ability, it is essential to add the penalty term to control the tendency of overfitting. L1 regularization is chosen which could make the parameter matrix sparse and lower the parameter numbers. And with good anti-interference ability against abnormal points, it is more robust\[21\].

\[ L = L_0 + \lambda \| \theta \| \]  

2.11. Converging condition

As the iteration increases, the fitting error will become decrease. At the same time, learning rate will also becomes smaller, which will cause the gradient to slow down or even disappear. In that case, training goes on endlessly, but the model is not optimized. So we need to set training termination conditions.

Under the mechanical performance test of steel material, the strength value of 0.1MPa and elongation of 0.1% are usually the limits of measurement accuracy. Therefore, it is appropriate to set the end of training error at 0.1. If the error cannot continue to decrease by 0.1 in 20 verification cycles, the training stops. Fix the model parameters to predict and evaluate.

3. Training and analysis

In this chapter, we will train and analyze different topologies of neural networks from three aspects: neurons numbers, hidden layer numbers, and the way of output.

3.1. Number of neurons

In order to clarify the impact of the number of neurons on the performance of the neural network, we use a single hidden layer, set the number of neurons to 10, 20, 30, 50, 100, 500, 1000, 10000. Then we import the best parameter combination to get 1000 performance in new data set, and compare with Table 2 to calculate the correct rate.

Figure 2. Effect of neuron numbers on model accuracy in single hidden layer network.
As we could see from Figure 1, the prediction accuracy improves significantly as the number of neurons adds. When it is from 10 to 100, the correct rates go up by more than 26%, especially for yield strength. But the promotion in accuracy adds slowly after neurons exceeds 100. When it is up to 10,000, the increase reaches only 1-3%. Looking at the accuracy growth of three performance indicators, except for the initial 10 neurons, the correct rate of yield strength is lower than tensile strength and elongation in other stages.

3.2. Number of layers
On the basis of section 3.1 above, we fix the neuron number to 100, and set hidden layer numbers as 1, 2, 3, 5, 10, 15, and 20, to study the effect of layer numbers on the performance.

![Figure 3](image)

Figure 3. Effect of hidden layer numbers on model accuracy in multi hidden layer network.

It can be seen from Figure 3 that as the depth of the network increases, the accuracy adds. After more than 3 layers, the improvement slows down. When the number of layers exceeds 10, the accuracy even decreases. Similarly, the prediction accuracy of yield strength is lower than the other two properties.

Why not try to use more neurons like 1000 or 10000? As we could see from Table 3, when there is only one hidden layer, the network parameters of 1000 and 10000 neurons are 10 and 100 times than 100 neurons. But when the hidden layer number is 3, this multiple soars to 90 and 8860, leading to a sharp increase in training difficulty. In general, in the case of models with same performance, we prefer to use the one with fewer parameters. There are two reasons, one is that fewer parameters will reduce the performance analysis time, especially on a large number of steel coils. Second, fewer parameters could greatly lower the risk of over-fitting. From another point of view, the model parameter numbers increase greatly, but the performance of the model is not improved. This is an overfitting phenomenon in itself.

![Table 3](image)

Table 3. Network parameters with different numbers of neurons.

| Neurons | One hidden layer | Three hidden layers |
|---------|------------------|---------------------|
| 100     | 2403             | 22,603              |
| 1000    | 24,003           | 2026,003            |
| 10000   | 240,003          | 200,260,003         |

3.3. Way of Output
For the structure of 3 layers and 100 neurons studied above, two different way with 1 or 3 output features are tested respectively. From Table 4, we are surprised to see that when there is only one output feature, the prediction accuracy of the model is generally higher than that three output features by 2.8-6.7% on three property indicators. After training and predicting 3 times respectively, the correct rates of Rp0.2, Rm and A80 have reached 74.%, 89.8% and 83.2%.

![Table 4](image)
Table 4. Correct rate under different outputs.

| Output feathers | Three outputs | One output |
|-----------------|---------------|------------|
| Rp0.2           | 71.6%         | 74.4%      |
| Rm              | 83.1%         | 89.8%      |
| A80             | 78.6%         | 83.2%      |

The reason why different outputs leading to different prediction accuracy, equation (6) and (7) can be used for analysis. When the model has 3 output features, the mean square error (MSE3) is average of sum of squared errors (MSE3). With the iterations increase, MSE3 keeps decreasing. However, during training, there may be a large error in one feather, while the other two are small. Although MSE3 remains decreasing, it cannot be guaranteed that the training error of each performance indicator decreases or not. When the output feature is only one, that's not going to happen. The training error always keeps a downward trend until the termination condition is triggered.

\[
MSE_3 = \frac{1}{3} \left( (y_{out1} - y_{true1})^2 + (y_{out2} - y_{true2})^2 + (y_{out3} - y_{true3})^2 \right)
\] (6)

\[
MSE_1 = (y_{out} - y_{true})^2
\] (7)

MSE3: Mean square error for 3 output features
MSE1: Mean square error for 1 output feature
y_{out}: Output value
y_{true}: True value

In general, the yield strength ratio of IF steel is generally about 0.4-0.6, but the results of section 3.1 and 3.2 above show that yield strength correct rate is much lower than tensile strength and elongation. Reasons for this result can be explained from two aspects. One is that the measurement of yield strength needs to adjust the value range of elastic modulus according to the tensile curve, which directly affects the value of yield strength. As we can see from the Figure 4, for the same engineering stress-strain curve, a small change in the line slope will cause the yield strength varying from 143 MPa to 147 MPa. While the tensile strength and elongation will not be similar. This will lead to inaccurate true values (labeled data) in model training, which also increases the prediction error.

On the other hand, the yield strength is extremely sensitive to flat elongation, far exceeding than chemical components or other processes. The model may not be able to accurately fit this feature among a large number of input features. The superposition of these two factors lead to a decrease in the accuracy of yield strength prediction.

![Figure 4. Typical IF steel engineering stress-strain curve.](image)

4. Discussion
It should be pointed out that this hyperparameter combination is only applicable to the specific data set in this paper. If the number of input or output features, training data set, etc. have changed, the parameter combination may not be suitable for the new model. And it should be retrained again.
4.1. Optimization
The training and debugging process of neural network is to reduce the variance and bias of prediction results continuously, as well as the impact of noise data. It usually requires hundreds of times or more. Although adding neuron and layer numbers are beneficial to improve the prediction accuracy, it will also increase training difficulty. It consumes a lot of computing power and time, but does not necessarily improve prediction performance. And for the optimization of deep neural networks, it is almost impossible for us to obtain the best of each parameter. What we can do is to use parameters as few as possible, and debug the model constantly to meet our demand.

An important aspect of model optimization is maintaining the validity of labeled data. Specifically, it is essential to keep the consistency of sampling locations and test methods. As we know, properties of steel coils at different positions vary greatly. If these sampling positions cannot be unified, materials with the same chemical composition and process may fluctuate, resulting in inaccurate data set and high training error. For the same reason, the uniformity of testing method and equipment should be consistent as well. Only by ensuring high quality of labeled data, can be possible to ensure that the model has good performance.

On the other hand, according to the research in section 3.3 above, in the case of multiple output features, the model accuracy using a single output is better. But obviously, it will greatly increase the number of parameters and models. In the next research, we should try to develop a new loss function suitable for multiple outputs, so that the model performance is not lower than the former, without adding parameters and training difficulty.

In recent years, a new kind of knowledge-based neural network has appeared, in which material thermodynamics and kinetics could be used to establish sub-models including element diffusion, precipitation, thermal deformation, austenite transformation, etc., together with production data, to build a unified model. In the future, a deep learning model combining with data and knowledge will be an important direction, and we will continue to carry out relevant research[22-24].

4.2. Dynamic model adjustment
In the continuous production process of automotive steel, especially for sheet products, it is necessary to establish a stable and versatile performance model, whether it is for product design, order matching, or improving product performance stability.

However, the fact is that production processes are full of variables. They may be from raw material, equipment and technical personnel operations, and are unpredictable and uncontrollable. In these cases, a static model cannot meet the demand. Therefore, it is strongly recommended to establish a dynamic model adjustment mechanism based on the latest production data, to adjust the model structure and parameters promptly. This usually includes two types of factors. First, a cold rolling production line has been put into operation for a long time, and the equipment is aging and wearing. The implementation of upgrades and modifications on it, resulting in major changes in the production line parameters before and after the transformation. Second, the state of production line is stable, but the source of raw materials or upstream production methods change. For example, for hot-rolled products, steel making using molten iron or scrap steel has a great impact on residual elements in material. And hot rolling process or performance changes will be inherited to the cold rolling process. These may cause the original model to deviate from the real-time status of production line, and reduce the accuracy of the model. Therefore, we need to use the production data in recent period to obtain a updating model in real time. This time cannot be too short, so as not to obtain insufficient data, leading to low generalization ability. It is better to use the latest 1-3 months as the adjustment cycle.

4.3. Unified model
The model in this paper is for IF steel prediction. As we all know, besides of the mild steel, there are many types of automotive steel, such as DP(dual phase steel), CP(complex phase steel), HSLA(high strength low alloy steel), BH(baking hardened steel) and so on. These steels have different composition system, processes, properties, production methods and application ranges. For example,
DP steel is mainly strengthened by martensite phase, which controls the ratio of martensite and ferrite to achieve different strength levels. HSLA is based on ferrite matrix and small proportion of pearlite, which achieves different strength by refining grains and controlling the size and distribution of Ti(C,N)x, Nb(C,N)x and other precipitates[25].

Of course, we hope to establish a unified model for all steel grades, by inputting any chemical composition and process parameters, we can get the performance data we need. We have already tried to use the multi-layer network in this paper as a general model, but the results show that prediction accuracy is far lower than the model trained by one single steel. To solve this problem, different steel grades could be classified by clustering algorithm first, then training under different classifications may achieve higher prediction accuracy. In the follow-up research, we will try to build a neural network model with clustering and regression functions.

5. Conclusion
(1) Within a certain range, increasing the number of neurons and layers could both improve the accuracy of the model. For a single hidden layer network, simply increasing neuron numbers has limited effect on the model performance. And when neural network depth reached a threshold, model performance may decrease.

(2) The prediction accuracy of yield strength is always lower than tensile strength and elongation, which has an important relationship with yield strength measurement method and flat elongation sensitivity.

(3) Due to the calculation difference of mean square error(MSE), the accuracy of a single output feature model is 2.8-6.7% higher than that of three outputs. And the prediction correct rate of yield strength, tensile strength and elongation could reach 74.4%, 89.8% and 83.2%.

(4) The model optimization direction should be started from the following aspects: regular model updating, labeled data expansion, continuous perparameters optimization, measurement method stabilization and adding knowledge-driven sub-model.

References
[1] Cowling, P. , & Rezig, W. . (2015). Integration of continuous caster and hot strip mill planning for steel production. Journal of Scheduling, 3, 4: 185-208.
[2] Sandberg, P. , Larsson, M. , Dahl, J. , Soederstroem, M. , & Vourinen, H. . (2006). In search of stability: investigating flexible and stable production strategies for an optimised steel plant. International Journal of Green Energy, 3, 2: 159-170.
[3] Cigizoglu, & Kerem, H. . (2005) Application of generalized regression neural networks to intermittent flow forecasting and estimation. Journal of Hydrologic Engineering, 10, 4: 336-341.
[4] Chasset, P. O. . (2013) Grnn: general regression neural network. Revue De Physique Appliquée, 6: 1321-1325.
[5] Goulermas, J. Y. , Liatsis, P. , Zeng, X. J. , & Cook, P. . (2007) Density-driven generalized regression neural networks (dd-grnn) for function approximation. IEEE Transactions on Neural Networks, 18, 6: 1683-1696.
[6] Hinton, G. E. , & Salakhutdinov, R. R. .(2006) Reducing the dimensionality of data with neural networks. Science, 313, (5786) :504-7.
[7] Funahashi K, Nakamura Y.(1993) Approximation of dynamical systems by continuous time recurrent neural networks, Neural networks, 6: 801-806.
[8] Hornik K, Stinchcombe M, White H.(1989) Multilayer feedforward networks are universal approximators, Neural networks,2: 359-366.
[9] Qiu X P. (2019) Neural Networks and Deep learning. Fudan University. https://nndl.github.io/nndl-book.pdf.
[10] Gulshan, V. , Peng, L. , Coram, M. , Stumpe, M. C. , Wu, D. , & Narayanaswamy, A. , et al. (2016) Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA, 316, 22: 2402-2410.

[11] Kohavi, R. . (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. In: International joint conference on Artificial intelligence. Morgan Kaufmann Publishers Inc. pp.1137–1143.

[12] Viana, F. , Haftka, R. T. , & Jr, V. S. . (2009) Multiple surrogates: how cross-validation errors can help us to obtain the best predictor. Structural & Multidisciplinary Optimization, 39, 4: 439-457.

[13] Saha, R. , & Ray, R. K. . (2008). Effects of severe cold rolling on the texture and grain boundary character of a Ti–Nb if steel. Materials Letters, 62, 2: 222-225.

[14] Chen, Y. L. , Wang, Y. , & Zhao, A. M. . (2012). Precipitation of AlN and MnS in low carbon aluminium-killed steel. Journal of Iron & Steel Research, 04: 51-56.

[15] M Yetna N’Jock, Chicot, D. , Decoopman, X. , Lesage, J. , Ndjak a, J. M. , & Pertuz, A. . (2013) Mechanical tensile properties by spherical macroindentation using an indentation strain-hardening exponent. International Journal of Mechanical Sciences, 75: 257-264.

[16] Schmidt-Hieber, J. . (2017) Nonparametric regression using deep neural networks with Relu activation function.

[17] Eckle, K. , & Schmidt-Hieber, J. . (2018) A comparison of deep networks with Relu activation function and linear spline-type methods. Neural Networks.

[18] Wang, P. . (2020). Based on adam optimization algorithm: neural network model for auto steel performance prediction. Journal of Physics Conference Series, 1653, 012012.

[19] Kingma, Diederik, Ba, Jimmy. (2015) Adam: A Method for Stochastic Optimization. In: Proceedings of the 3rd International Conference on Learning Representations, ICLR.

[20] Ding C C, Tao W, Tao Q. (2020) A unified momentum method with triple-parameters and its optimal convergence rate, Journal of computer research and development, 57: 1571-1580.

[21] Zou H, Hastie T.(2005) Regularization and variable selection via the elastic net, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67:301-320.

[22] Karpatne, A. , Atluri, G. , Faghmous, J. H. , Steinbach, M. , Banerjee, A. , & Ganguly, A. , et al. (2017). Theory-guided data science: a new paradigm for scientific discovery from data. In: IEEE Transactions on Knowledge & Data Engineering, PP.99.

[23] Yaman, B. , Hosseini, S. , Moeller, S. , Ellermann, J. , Uurbil, K. , & Akakaya, M. . (2020). Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data. Magnetic Resonance in Medicine, 84, 6: 3172-3191.

[24] Wang, L. , Zhou, Q. , & Jin, S. . (2020). Physics-guided deep learning for power system state estimation. Journal of Modern Power Systems and Clean Energy, 8,4: 607-615.

[25] Wang, J. Y. , Liu, J. H. , Liu, J. F. , Feng, M. L. , & Jia, Y. S. . (2014). Precipitation mechanism and behavior of tin during Ti-If steel solidification. Journal of University of Science and Technology Beijing, 36, 8: 1025-1031.