한국기계가공학회지, 제21권, 제6호, pp. 89～97(2022.05) ISSN 1598-6721(Print)
Journal of the Korean Society of Manufacturing Process Engineers, Vol. 21, No. 6, pp. 89～97(2022.06) ISSN 2288-0771(Online)

https://doi.org/10.14775/ksmpe.2022.21.06.089

In-process Weld Quality Monitoring by the Multi-layer Perceptron Neural Network in Ultrasonic Metal Welding

Muhammad Bilal Shahid*, Dong-Sam Park**,#
*Department of Mechanical Engineering, Iowa State University
**Department of Mechanical Engineering, Incheon National University

초음파 금속용접 시 다층 퍼셉트론 뉴럴 네트워크를 이용한 용접품질의 In-process 모니터링

*Muhammad Bilal Shahid, 박동삼**,#
*아이오와 주립대학교 기계공학과, **인천대학교 기계공학과

(Received 25 April 2022; received in revised form 12 May 2022; accepted 19 May 2022)

ABSTRACT

Ultrasonic metal welding has been widely used for joining lithium-ion battery tabs. Weld quality monitoring has been an important issue in lithium-ion battery manufacturing. This study focuses on the weld quality monitoring in ultrasonic metal welding with the longitudinal-torsional vibration mode horn developed newly. As the quality of ultrasonic welding depends on welding parameters like pressure, time, and amplitude, the suitable values of these parameters were selected for experimentation. The welds were tested via tensile testing machine and weld strengths were investigated. The dataset collected for performance test was used to train the multi-layer perceptron neural network. The three layer neural network was used for the study and the optimum number of neurons in the first and second hidden layers were selected based on performances of each models. The best models were selected for the horn and then tested to see their performances on an unseen dataset. The neural network models for the longitudinal-torsional mode horn attained test accuracy of 90%. This result implies that proposed models has potential for the weld quality monitoring.

Keywords : Weld Quality(용접 품질), In-process Monitoring(인프로세스 감시), L-T horn(L-T 혼)Ultrasonic Welding(초음파 용접), Neural Network(뉴럴 네트워크)

1. Introduction

Ultrasonic welding (USW) is remarkable application of ultrasonic in engineering where an ultrasonic weld horn vibrating at a high frequency(i.e. 20~40 kHz) is used to join the two or multiple sheets of metal foils[1]. The main component of a USW machine is the weld horn which is also the focus of this paper.

There are different types of a USW weld horn

Corresponding Author : dspark@inu.ac.kr
Tel: +82-32-835-8418, Fax: +82-32-835-0793

(i.e., longitudinal(L) and longitudinal-torsional(L-T)) which depend on their vibration modes. There are different ways of achieving composite mode in weld horns. It can either be achieved by introduction of the slanting grooves to the front part of the horn[2] or by using coupled transducers where two sets of piezoelectric polarized in thickness and tangential directions are used[3]. In a previous study, high strength welds were formed with a comparatively shorter welding times and smaller vibration velocities when using L-T mode in USW[4,5].

Another important topic in USW is weld quality monitoring which is also addressed in the present work via neural network(NN) approach. The quality of an ultrasonic weld depends on the intensity of welding parameters like pressure, amplitude, and time[6] and it turns out that the suitable values of these parameters must be chosen to obtain an optimum quality weld[7]. There are many approaches to weld quality monitoring that were already introduced by others. One such approach is the use of dynamic resistance measured during resistance spot welding and NN for the quality prediction[8]. In another study based on thin-plate laser welding, NN optimized by the genetic algorithm and principal component analysis was used for real-time weld geometry prediction[9]. Another real-time weld monitoring algorithm was developed for the disk laser welding where some features were selected manually as an input to the classical machine learning algorithm(ML) algorithm support vector machines.

Although there are many studies already focused on the performance of L-T modes ultrasonic devices in different contexts such as cutting and drilling, the studies based on the performance of L-T modes particularly in the context of the USW are very rare. Moreover, this study uses the multi-layer perceptron (MLP) neural network to predict the strength of the welds. This MLP which takes as input the values of USW parameters predicts the weld strength and has the potential to be used in USW quality monitoring. This simplest and fast NN is very helpful for USW monitoring given that USW is characterized by the shorter welding times typically on the order of milliseconds.

2. Overview of MLP

As discussed in the earlier section, there are many approaches to weld monitoring involving the use of different types of ML and genetic algorithms but not many researchers are interested in harnessing the power of MLP for weld monitoring. This simple yet so useful ML algorithm is very fast compared to deep neural networks such as convolutional neural network(CNN) and a good fit for the monitoring purposes.

2.1 Perceptron

In MLP, the perceptron is the most basic unit as shown in the Fig. 1 and has been referred to as the black box by many researchers. In essence, it’s a simple approximation that takes some inputs x, multiplies them with their weights w, add a bias b, and gives us output y after introduction of non-linearity via an activation function σ. It is given as:

$$z = wx + b$$
$$a = \sigma(z) = y$$

The weights in perceptron are the parameters that indicated the importance of each input value x. These weights keep on changing their values as we train the neural network. If enough training data are provided, a perceptron can approximate the function very well.

In MLP, many perceptron are stacked above each other and many layers are stacked one after another. By adding many layers, the complex problems can be approximated via training. At first, input layer holds all the inputs and these inputs are fed to the hidden layer which then performs some approximations as stated above. The outputs of this layer are then fed to output layer as shown in the Fig. 2 below. Such a network in which each input from the previous layer is passed to each perceptron neuron) in the next layer is called densely connected NN. This simple MLP is called a shallow NN. One thing to note is that the input layer is usually ignored when counting the number of layers in MLP. As such, the network in Fig. 2 is a two layer network and the MLP is named as deep NN if it has three or more layers.

As can be seen in Fig. 2, the activations of each individual layer are denoted by the letter a where the subscript of the ‘a’ indicates the position of each neuron within each layer and the superscript indicates the number of the layer to which the neuron belongs. For example, the activation shows that it is the third neuron in the first hidden layer. The notations for weights and other parameters can be interpreted in the same way.

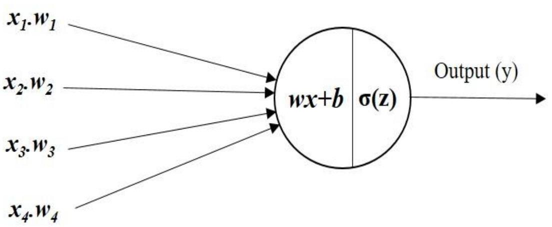


Fig. 1 The fundamental unit of MLP

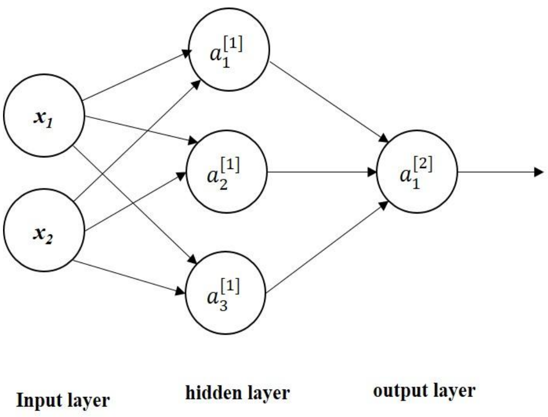


Fig. 2 A shallow two layer MLP

2.2 Activation function

These are the functions that are used in neural networks and that computes the weighted sum of input and biases which then determines whether a specific neuron should be activated or not[10]. They can be linear or non-linear depending on the problem under consideration but the usage of non-linear activation functions(AFs) is more common since it makes the neural network to learn the parameters in an easier way. There are three most common AFs that are being used.

2.2.1 Sigmoid function

It is a commonly used AF and also referred to as the logistic function. Although there are many variants of it such as hard sigmoid function, sigmoid-weighted linear units(SiLU), and derivative of sigmoid-weighted linear(dSiLU), the simplest one will be discussed here. This AF is used in feedforward neural network and it is differential with positive derivatives everywhere[11]. It is given as:

$$f(x)=\frac{1}{(1+e^{-x})}$$

This AF is mostly used for outputs of the neural networks in the form of probabilities such as binary classification problems. The advantage of this AF is that they are very useful to model logistic regression tasks and they are used mostly in shallow neural networks[12].

2.2.2 Hyperbolic tangent function (tanh)

It is another function that is used in deep neural networks and has some variants like sigmoid function. This AF is zero-centered (i.e., centers all the values on zero) with all the values lie between -1 and 1[13]. It is given as:

$$f(x)=\frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}$$

This AF gives better performance for multi-layer networks[14] and being a zero-centered function it aids in the back-propagation process. It is mostly used in speech recognition and natural language processing tasks.

2.2.3 Rectified linear unit (ReLU)

This AF has been the most popular one for deep architectures and it has achieved state-of-the-art results for many deep learning tasks[15]. It learns faster and is easier to optimize with gradient descent methods thus performs better than Sigmoid and tanh functions. Unlike previous two AFs, this function does not need to compute exponentials and addition/multiplication operations. It is a simple threshold function that sets all the values less than zero to zero unlike sigmoid and tanh AFs thus avoiding vanishing gradient problems. It is given as:

$$f(x) = ax + x = \begin{cases} x & \text{if } x > 0 \\ ax & \text{if } x \le 0 \end{cases}$$

Due to its popularity, faster computation, better optimization, and excellent performance, it was used in the current study.

3. Principle of ultrasonic metal welding

Fig. 3 shows the schematics of ultrasonic metal welding. The converter, booster, and horn convert electrical energy to mechanical vibrations of 20kHz ~40kHz. The horn transfers the mechanical vibration energy to the weld materials finally. When the amplified ultrasonic vibration is transmitted to the weld through the horn, strong joining is achieved by solid-state diffusion in the weld interfaces.

In the ultrasonic metal welding, welding parameters are welding energy, clamping pressure, welding time, and amplitude of horn vibration. These parameters have great effects on the weld quality, and then should be appropriately selected. The welding energy delivered by the machine is directly proportional to pressure, amplitude, and time.

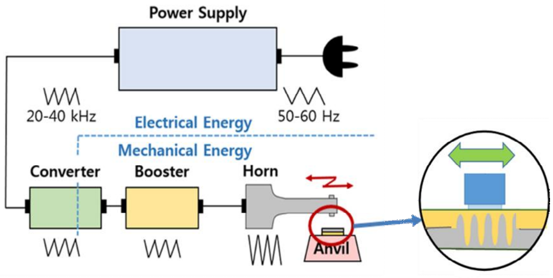


Fig. 3 Schematics of ultrasonic metal welding

4. Experiments

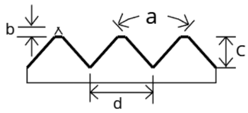
4.1 Specimen and welding machine

Two sheets of 0.1mm thick Cu and Ni are welded. The specimen was 10mm wide and 50mm long. The welding machine is D9800, which is made in DURASONIC Co. LTD., Korea.

4.2 Ultrasonic horn

The ultrasonic horn plays a key part in ultrasonic welding. In our previous study[2], a composite mode horn with the resonance of 20kHz for ultrasonic welding was developed by introducing slanting grooves at the front mass of the horn. The L-T mode horn used in experiment is shown in the Fig. 4(a). The knurling pattern on the horn tip is shown in the Fig. 4(b).

(a) Shape of L-T mode horn



a	b	c	d
90⁰	0.05	0.10	0.30

(b) Knurling pattern on the horn tip(mm)

Fig. 4 L-T mode horn and knurling pattern[2]

4.3 Welding method

The L-T horn vibrates in both longitudinal and torsional directions. USW quality depends on many weld parameters such as pressure, amplitude, and time[6]. Three different levels of these three parameters were chosen for experimentation and the specific values of these parameters are shown in the Table 1.

Table 1 Welding parameters used in the experiment

Welding parameters	Levels
Pressure(MPa)	0.1, 0.2, 0.3
Amplitude(μm)	16, 24, 32
Time(sec)	0.1, 0.2, 0.3

4.4 Welding strength measurement method

To measure welding strength of welded specimen, U-tensile tests were made using COMETECH® tensile testing machine with a 1kN load cell.

As described previously, seven experiments were made against each welding condition. Their strength results were then averaged to get the mean strength as shown in Table 2. As can be seen, the strength increases with the increase in value of weld parameters. For example, the first three rows indicate that the strength increase with respect to increase in value of time while keeping two other parameters constant and the same is true for pressure and amplitude as well.

5. Weld quality monitoring approach

As explained earlier, the NNs are very good at function approximation(i.e. mapping functions from some input to an output). In the current study, the data collected for the performance test of L-T mode horn will be used for training and testing of the MLP. This trained MLP will have the potential of the USW monitoring. As the strength is an indicator of the weld quality, this trained NN will predict the strength by taking the weld parameters as its inputs.

Table 2 The mean strengths of the welds

Pressure (MPa)	Amplitude (μm)	Time (sec)	Welding strength (N)
0.1	16	0.1	15
0.1	16	0.2	25
0.1	16	0.3	26
0.1	24	0.1	21
0.1	24	0.2	25
0.1	24	0.3	27
0.1	32	0.1	25
0.1	32	0.2	29
0.1	32	0.3	30
0.2	16	0.1	10
0.2	16	0.2	11
0.2	16	0.3	24
0.2	24	0.1	25
0.2	24	0.2	27
0.2	24	0.3	29
0.2	32	0.1	25
0.2	32	0.2	29
0.2	32	0.3	30
0.3	16	0.1	9
0.3	16	0.2	24
0.3	16	0.3	23
0.3	24	0.1	20
0.3	24	0.2	22
0.3	24	0.3	26
0.3	32	0.1	24
0.3	32	0.2	30
0.3	32	0.3	32

5.1 Training and testing of MLP

The NN with one input, one output, and two hidden layers was trained using the data collected for performance test. By keeping the number of layers fixed, there is still some tuning required to find the optimized number of hidden neurons for both first and second hidden layers. Many different architectures were tried with varying configurations and the best among those was selected based on the training and validation accuracies. The different number of neurons were selected for hidden layers as shown in Table 3.

Data were divided into training (70%), validation (15%) and testing (15%). The purpose of the validation data is to validate the accuracy of various models after training so that the best configuration can be selected. The testing data should then be used to test the best configuration performance.

Table 3 Number of neurons against two hidden layers

Layer no.	No. of neurons
First hidden layer	10, 15, 20, 25, 30
Second hidden layer	5, 10, 15

The validation data cannot be used to test the accuracy of the best configuration since that data had already been seen by the NN and thus unable to give an unbiased performance indication. The different models resulted from the different number of neurons were trained using the training data and the performance was validated using the validation data.

Before feeding the data to the neural network, the normalization was performed using the StandardScalar from Scikit-learn[16]. This is required because the input features are usually on a different scale and normalization transforms the data to have the mean of 0 and standard deviation of 1. To compile the model, mean squared error, Adam, and coefficient of determination(R^2) were used as the loss, optimizer, and the metrics of accuracy respectively. The coefficient of deviation(R^2) used to determine the model accuracy can be given as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n-1} (y_i - p_i)^2}{\sum_{i=1}^{n-1} (y_i - \mu_{test})^2}$$

where the labeled true output value and the predicted value of each dataset instance i are denoted by y_i and p_i respectively, while denotes the mean value of the data set. The mean squared error can be written as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - p_i)^2$$

where the labeled true output value and the predicted value of each dataset instance i are denoted by p_i and y_i respectively. These data can be passed to the neural network in two different ways i.e., gradient descent and mini-batch gradient decent. The batch gradient descent updates the weights only once all of the data has been passed through the neural network while mini-batch gradient decent passes the data in small batches(e.g. 32, 64, 128 etc.) and updates the weights once that small batch has been passed through the neural network. Thus, the mini-batch gradient descent results in faster weight updates and more efficient compared to batch gradient descent. In the current study, the data was passed in the batches of 64.

The data for the L-T mode horn was used to train many different models resulted from the different number of neurons in the first and second hidden layers. The performances of different models were compared based on the training and validation accuracies. The number of neurons in the first and second hidden layers of the best model are 20 and 10 respectively and its architecture is shown in the Fig. 5. The best model obtained had the training and validation accuracies of 93 % and 91 % respectively. The trends of training and validation accuracies of the model with respect to the number of epochs are shown in Fig. 6 while the loss in given in Fig. 7. The model had some problems learning from the training data as can been from the noisy training accuracies in Fig. 6 but it actually gets stable by the end of training (i.e. 800 epochs). On the other hand, the model validation accuracy although remained lower than the training accuracy was stable throughout the validation process.

The testing data was used to test the accuracy of the best model. The model performed reasonably well with an accuracy of 90% as shown in the Fig. 8.

The model predictions were very close to the target output values.

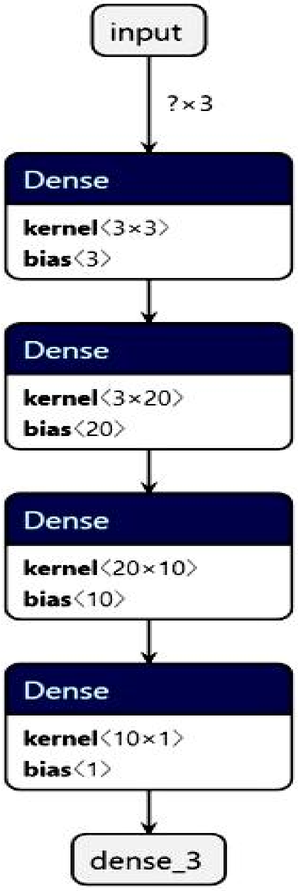


Fig. 5 The best architecture for L-T horn

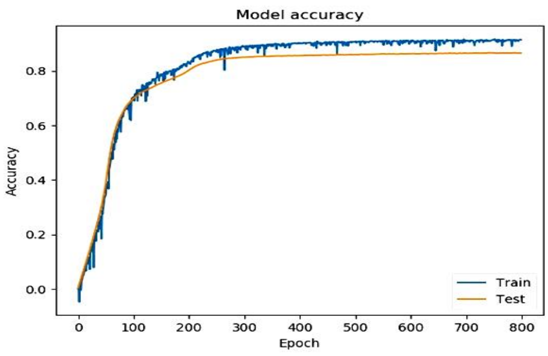


Fig. 6 Training and validation performance

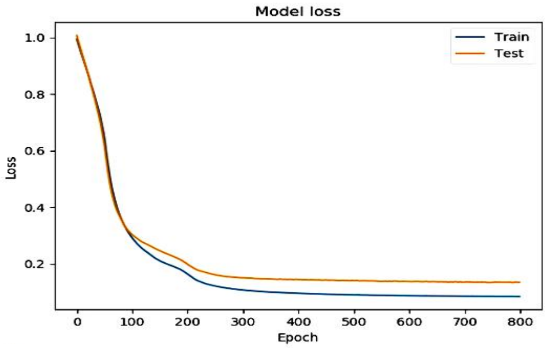


Fig. 7 Loss during training and validation

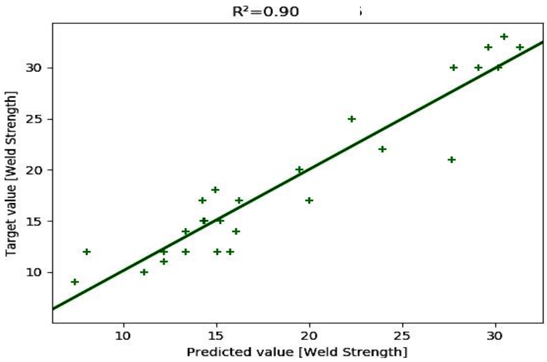


Fig. 8 The predicted values vs the true output labels

5.2 Algorithm description

As stated earlier, the USW quality depends on the weld parameters such as pressure, time, and amplitude. The NN was trained and tested for L-T mode horn. The trained neural network has the potential to be implemented in a USW machine. The user needs to specify the value of these three parameter before the machine starts making welds. These parameters could be input to the NN model automatically which then predicts the weld strength corresponding to the weld parameters. As the weld strength is an indicator of the weld quality, the machine would classify the welds as good/bad depending on the predicted strength. The strength range for good/bad welds needs to be specified by the user as it varies depending on the thickness and weld material.

5. Conclusion

The MLP neural network was trained using the weld data that was collected for the performance test. In case of L-T mode horn, the training and validation accuracies for the best model were 93 and 91 percent respectively with 20 and 10 as the optimum number of neurons for the first and second hidden layers respectively. These best performing models were tested on a separate test dataset. The MLP trained on the L-T mode dataset attained 90 %. The NN models for both the horns have the potential to predict the weld strengths and be implemented in a USW machine for weld quality monitoring.

The MLP model trained on L-T mode horn data attained an accuracy of 90 %. The NN models for the L-T mode horn have the potential to predict the weld strengths and be implemented in a USW machine for weld quality monitoring.

Acknowledgement

This work was supported by Incheon National University Research Grant in 2020.

References

1. Shahid, M. B., Han. S.-C., Jun, T. S., Park, D. S., “Effect of process parameters on the joint strength in ultrasonic welding of Cu and Ni foils,” Materials Manufacturing Processes, Vol. 34, pp. 1217-1224, 2019.
2. Shahid, M. B., Jung, J. Y., Park, D. S., “Finite element analysis coupled artificial neural network approach to design the longitudinal-torsional mode ultrasonic welding horn,” International Journal of Advanced Manufacturing Technology, Vol. 107, pp. 2731-2743, 2020.
3. Lin, S., “Sandwiched piezoelectric ultrasonic transducers of longitudinal-torsional compound vibrational modes,” IEEE Transaction on Ultrasonics, Ferroelectrics, and Frequency Control, Vol. 44, No. 6, pp. 1189-1197, 1997.
4. Tsujino, J., Ueoka, T., Hasegawa, K., et al., “New Methods of Ultrasonic Welding of Metal and Plastic Materials,” Ultrasonics, Vol. 34, pp. 177-185, 1996.
5. Tsujino, J., “Recent Developments of Ultrasonic Welding,” Proc. of 1995 IEEE Ultrasonics Symposium, pp. 1051-1060, 1995.
6. Lee, S. S., Kim, T. H., Hu, S. J., Cai, W. W., Abell, J. A., Li, J., “Characterization of Joint Quality in Ultrasonic Welding of Battery Tabs,” Journal of Manufacturing Science and Engineering, Vol. 135, No. 2, pp. 021004, 2013.
7. Chang, U. I., Frisch, J., “On optimization of some parameters in ultrasonic metal welding,” Welding Journal, Vol. 1, pp. 24-35, 1974.
8. Wan, X., Wang, Y., Zhao, D., Huang, Y., Yin, Z., “Weld quality monitoring research in small scale resistance spot welding by dynamic resistance and neural network,” Measurement, Vol. 99, pp. 120-127, 2017.
9. Lei, Z., Shen, J., Wang, Q., Chen, Y.,“Real-time weld geometry prediction based on multi- information using neural network optimized by PCA and GA during thin-plate laser welding,” Journal of Manufacturing Processes, Vol. 43, pp. 207-217, 2019.
10. Nwankpa, C. E., Ijomah, W. I., Gachagan, A., Marshall, S., “Activation Functions: Comparison of Trends in Practice and Research for Deep Learning,” 2nd International Conf. on INCCST, Jamshoro, Pakistan, 2020.
11. Han, J., Moraga, C., “The influence of the sigmoid function parameters on the speed of backpropagation learning,” Lecture Notes in Computer Science, Springer, pp. 195-201, 1995.
12. Neal, R. M., “Connectionist learning of belief networks,” Artificial Intelligence, Vol. 56, pp. 71-113, 1992.
13. LeCun, Y., Bengio, Y., Hinton, G., “Deep learning,” Nature, Vol. 521, pp. 436-444, 2015.
14. Karlik, B., Olgac, A. V., “Performance Analysis of

Various Activation Functions in Generalized MLP Architectures of Neural Networks," International Journal of Artificial Intelligence And Expert Systems(IJAE), Vol. 1, Issue 4, pp. 111-122, 2011.
15. Nair, V., Hinton, G. E., "Rectified Linear Units Improve Restricted Boltzmann Machines," Proc. of the 27th International Conference on Machine Learning, Israel, pp. 807-814, 2010.
16. Pedregosa, F., Varoquaux, G., Gramfort, A., et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, Vol. 12, pp. 2825-2830, 2011.