Improving of Classification Accuracy based on Stacked autoencoder Feature extracted via Bayesian Optimization

J. E. Song[1] and S. J. Bae[1*]

[1]Department of industrial Engineering, Hanyang University, Seoul, Republic of Korea

[*]Corresponding author: sjbae@hanyang.ac.kr

1. Introduction

This study proposes to extract the characteristics of data measured from mechanical components when diagnosing a failure of a machine and to make a quick diagnosis on the failure of a part. In particular, since bearings are a key component of machine, they are traditionally diagnosed through indicators such as health index, Fast Fourier Transform to determine bearing failure.

In this study, we used the Stacked Auto Encoder as a model to extract the features of the vibration signal measured using an accelerometer. In the process, Bayesian optimization will be used to improve the performance of the model. After the optimum value was obtained through Bayesian optimization, the extracted features were placed in the classifier to improve the accuracy of classification of normal and fault based on the degradation starting point. As a result, the classification performance was higher when Bayesian optimization was used to optimize the hyper-parameter.

2. Bayesian optimization

Bayesian optimization is a derivative-free method for global optimization. Derivative-free optimization doesn't use derivative information in the classical sense to find optimal solutions. Instead, Bayesian optimization is inspired by the Bayesian method of updating a given model of observations. This method focused on solving the following problem

$$x^* = argmin_{x \in A} f(x). \tag{1}$$

Where $f(x): A \rightarrow \mathbb{R}$ is called the objective function[1].
Eq. (1) have equally mathematically formulated the problem as minimization since

$$x^* = argmin_{x \in A} f(x) = argmax_{x \in A} - f(x). \tag{2}$$

Bayesian optimization consists of two important elements. The first is a Bayesian statistical model of the objective function, and the second is and acquisition function that determines where to find the next sampling point. The underlying probabilistic model of the objective function f will typically be required as Gaussian process regression. Gaussian process makes it possible to obtain a predicted mean, which is considered as the potential optimal value, and the predicted standard deviation, which is considered as the potential risk, on any point. Therefore, we need a policy on which sample point to detect next. This policy is called the acquisition function. The key property that an acquisition function has to satisfy is a trade-off between exploitation and exploration, where exploitation concerns preferring the points with higher posterior mean, while exploration concerns preferring points with higher posterior variance[2]. When referring to acquisition functions, typical methods include probability of improvement(PI), expected improvement(EI).

3. Stacked auto-encoder

Stacked auto-encoder(SAE)[3], mainly used in the field of machine learning, features extracting important characteristics of data while reducing the size of existing data.

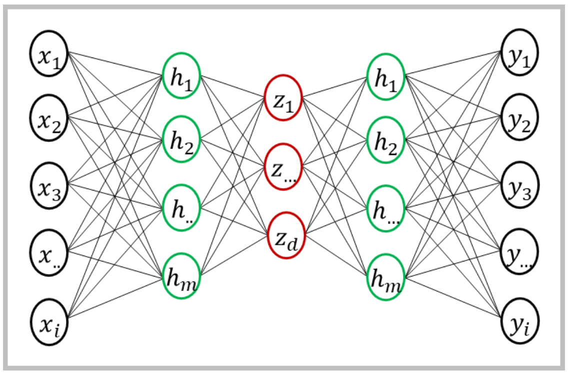

Fig.1 Stacked auto-encoder architecture

Since SAE is an unsupervised learning algorithm, training example is not labeled. In Fig.1 The characteristic of SAE is that the input value x_i and output value y_i are the same, so the whole process of auto encoding is to compare this reconstructed input to the original and try to minimize this error to make the reconstructed value as close as possible to the original[4]. Also, hidden nodes z_d can be used to extract the features of the input data.

4. Simulations and Result

The IMS bearing dataset used for this experiment has been collected on an endurance test rig of the

University of Cincinnati and released in 2014. An AC motor keeps the rotation speed constant. The four bearing are in the same shaft and are forced lubricated by a circulation system that regulates the flow and the temperature. Datasets contain 1-second recordings from the accelerometers with a sampling frequency of 20KHz, every 10min, during the run-to-failure tests[5]. The following table 1 is the information about the dataset2 used in the IMS bearing dataset.

Table 1 Dataset description

	Dataset 2
Number of samples	984
Sample size per bearing	20480
Faulty Bearing	Bearing 1
Endurance duration	2004/02/12 ~ 2004/02/19 6days 20h (9840min)

Using dataset2 as input data to the SAE model, feature extraction was performed on the failed bearing1. Fig.2 below is the overall procedure of our experiment.

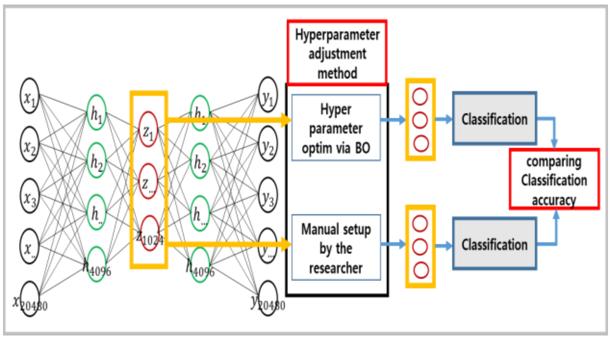

Fig.2 Simulation procedure

The cost function $J(\theta)$, an indicator of the performance of the SAE model, is influenced by the weight setting, and the Adaptive Moment Estimation(ADAM) is used to update the these weight. When updating weights using the ADAM method, the learning rate η, beta1 β_1, beta2 β_2, and epsilon ϵ exist as hyper-parameters. In the experiment, we fixed β_1, β_2, and ϵ at 0.9, 0.999, and 10^{-8} respectively, and performed Bayesian optimization for the learning rate η.

As shown in Fig.2, we compared the classification performance of hyper-parameter adjustment with Bayesian optimization and manual adjustment. After each feature extraction, normal and fault labeling was performed on the basis of the degradation starting point[6] and then classified using the deep neural network.

As a result, we confirmed that the feature where hyper-parameter was adjusted by Bayesian optimization had higher classification performance. Through this, we showed that when extracting the feature of data using the SAE, the Bayesian optimization was applied to improve the classification performance by better reflecting the feature of data.

References

[1] FRAZIER, Peter I. A tutorial on bayesian optimization. *arXiv preprint* arXiv:1807.02811, (2018)

[2] MATOSEVIC, Antonio. On Bayesian optimization and its application to hyperparameter tuning. (2018).

[3] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. *Science*, (2006), 313(5786):504-507.

[4] TAO, Siqin, et al. Bearing fault diagnosis method based on stacked autoencoder and softmax regression. In: *2015 34th Chinese Control Conference* (CCC). IEEE, 2015. p. 6331-6335.

[5] Hai Qiu, Jay Lee, Jing Lin, and Gang Yu. Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics. *Journal of sound and vibration*, 289(4):1066–1090, (2006).

[6] Seokgoo Kim, Sungho Park, Ju-Won Kim, Junghwa Han, Dawn An, Nam Ho Kim, and Joo-Ho Choi. A new prognostics approach for bearing based on entropy decrease and comparison with existing methods. *Annual Conference of the Prognostics and Health management Society*, 2016.