Fully Unsupervised Defect Clustering using Adversarial Autoencoder and Bayesian Mixture Model

Taewan Kim[1] and Seungchul Lee[2]

[1,2]*Department of Mechanical Engineering, Pohang University of Science and Technology, Pohang, Gyeongsangbuk-do, 37673, Republic of Korea*

twkim97@postech.ac.kr
seunglee@postech.ac.kr

ABSTRACT

Identifying defect types and developing proper maintenance strategies is a major concern in modern industry. Most conventional studies have been conducted primarily based on a supervised learning scheme. However, supervised learning has a critical limitation in that it requires labeled data, which is difficult and expensive to obtain in real-world industry. Considering that there are many industries that do not perform post investigations on the defects, fully unsupervised learning methods, which do not exploit any information such as label data or the number of types, need to be developed. Accordingly, in this study, we propose a fully unsupervised defect clustering method that does not exploit any information other than the data itself. The proposed method consists of two major components. The first is dimensionality reduction into latent space via adversarial autoencoder, and the second is a Bayesian mixture model for distribution estimation in latent space. The experiments on a rolling-element-bearing dataset validate the effectiveness of our method. Specifically, our method performs defect clustering without any information other than the data itself, which is promising for real industrial applications.

1. INTRODUCTION

In modern industry, the importance of defect detection cannot be overstated. Defects in processes can lead to a variety of negative consequences, including increased costs, reduced productivity, and potential safety hazards. Therefore, it is critical to have effective defect detection methods in place to ensure that products meet quality standards and are safe for use.

In recent years, there has been a growing interest in using machine learning and deep learning techniques for defect detection in the industry. (Hoang, 2019, Duan, 2018) These techniques have the potential to identify and classify defects in real-time automatically. One common approach is to use acoustic and vibration data to detect defects in machinery. In this case, machine learning models are trained on time-series data to identify patterns that indicate a potential defect. Deep learning techniques, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have been shown to have promising results in defect detection due to their ability to learn time-domain or frequency-domain features from the data automatically. (Liu, 2016, Liu, 2018)

However, most of the conventional deep learning-based defect detection methods are based on a supervised learning scheme. Since supervised learning relies heavily on labeled data, which can be time-consuming and expensive to obtain in industrial, it is impractical in real-world industrial applications. Moreover, supervised models are limited by the specific types of defects they are trained to detect and may not be able to detect previously unseen defects. In contrast, unsupervised learning techniques have the potential to detect novel and complex defects without the need for labeled data, making them more adaptable and versatile in real-world industrial applications.

In this study, we propose a fully unsupervised defect clustering method that can overcome the limitation of supervised methods. The proposed method does not exploit any information other than the data itself. Our method estimates the total number of defect types from the data and clusters them simultaneously, which is promising for analyzing large-scale data where it is difficult to obtain exact information about the defect types. The effectiveness of this method was validated in experiments on a rolling-element bearings defect dataset.

2. PROPOSED METHOD

2.1. Adversarial Autoencoder

Adversarial Autoencoder (AAE) is a type of deep learning algorithm that combines generative models and adversarial training to learn useful data representations in an unsupervised manner (Makhzani, 2015). The AAE model consists of two components: an encoder network that maps input data to a latent space and a decoder network that generates output data from the learned latent representation. However, unlike traditional autoencoders, AAEs introduce an adversarial component (discriminator) that is trained to distinguish between real and fake data samples in the latent space. The discriminator is trained to discriminate whether the samples are from the encoder or the prior distribution. This adversarial training improves the quality of the latent representation, making it more robust to variations in the input data and better at preserving important features. In this study, we designed the AAE model in a similar structure to the previous study (Kim, 2022). We preprocess the time-domain vibration signal to the frequency-domain data using a fast Fourier transform (FFT), and the AAE encodes the FFT results to the latent space.

2.2. Bayesian Mixture Model

The Bayesian mixture model is a statistical model that assumes a dataset is generated from a mixture of several subpopulations or clusters, each represented by a different probability distribution. This model uses a Bayesian approach to estimate the parameters of the probability distributions and the number of subpopulations. Additionally, the number of subpopulations is not fixed and is treated as a random variable determined by the data. In this study, we applied the Bayesian mixture model to the latent variables to estimate the number of clusters while simultaneously clustering the subpopulations.

2.3. Dirichlet Process and Prior Distribution Estimation

An AAE trains the encoder so that the encoded latent variables follow a pre-determined prior distribution. However, since we have no explicit information about the prior distribution, we need to set the prior distribution arbitrarily. In this study, we estimate the prior distribution via a Dirichlet process (DP). DP is a stochastic process whose range is itself a set of probability distributions. In this study, we estimate the mixture of Gaussian by DP on the encoded latent variables. Based on the estimated Gaussian mixture distribution, adversarial loss enforces the encoder to map the input to the Gaussian mixture prior. We train the autoencoder, estimate the Gaussian mixture by DP, and train AAE, alternately. As a result, the proposed AAE can estimate the proper number of clusters on its own for an arbitrary dataset without any information and perform clustering accordingly. A brief illustration of the proposed method is shown in Figure 1.

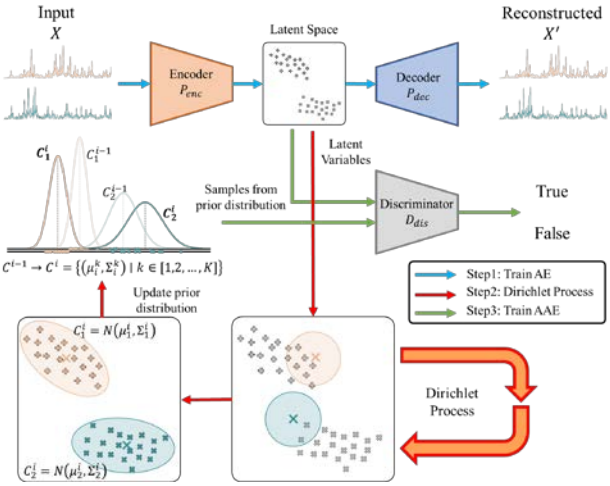

Figure 1. Illustration of the overall AAE training scheme.

3. RESULTS AND DISCUSSIONS

3.1. Experiment Setting

The effectiveness of the proposed fully unsupervised clustering method was validated using an experimental bearing dataset from the Case Western Reserve University (CWRU) in the United States. The bearings in the dataset were artificially damaged in different locations and with different severities, as listed in Table 1. Even if the defects were in the same location, the vibration patterns varied depending on the damage method and severity. Therefore, both damage type and severity were considered for class labeling in accordance with the problem definition used in a previous study (Li, 2020). We set detailed experimental settings by referring to a previous study (Kim, 2022).

Table 1. Labels and fault descriptions of the CWRU dataset. N is a normal condition bearing, REF is a rolling element fault, IF is an inner race fault, and OF is an outer race fault.

Class label	1	2	3	4	5
Fault type	N	REF	REF	REF	IF
Fault diameter (inch)	-	0.007	0.014	0.021	0.007
Class label	6	7	8	9	10
Fault type	IF	IF	OF	OF	OF
Fault diameter (inch)	0.014	0.021	0.007	0.014	0.021

3.2. Results

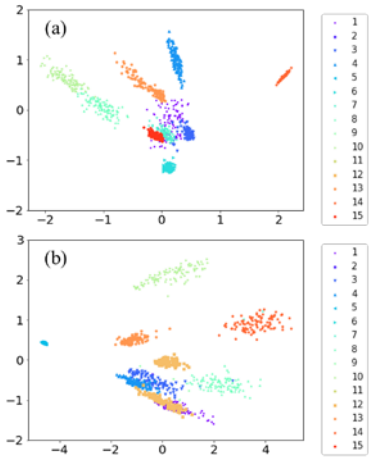

Figure 2. Illustration of the latent representation (2D, PCA) of the inputs. Legend means cluster index. (a) is a result of latent variables encoded by AAE with DP and (b) is a result of latent variables encoded by autoencoder.

First, we compared the proposed DP-integrated AAE with a naive autoencoder, and the latent representations of the inputs are illustrated in Figure 2. The proposed DP-integrated AAE showed a more distinct separation in the latent space representation than the naive autoencoder. This result demonstrates that the proposed AAE model maps the input data to the latent space in a way that accentuates the differences between the data.

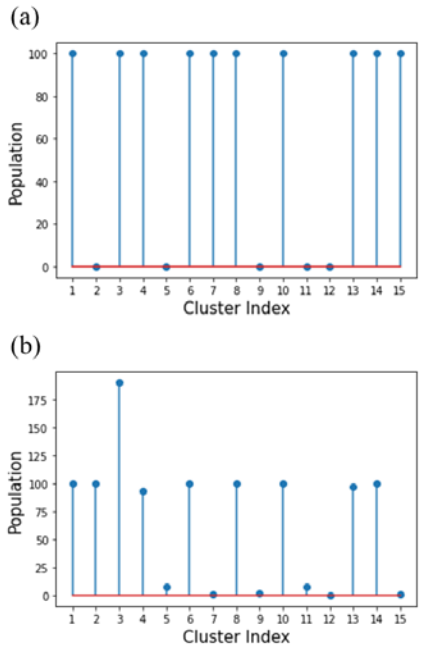

Figure 3. Unsupervised clustering results of Bayesian mixture model. (a) is a clustering result on the DP-integrated AAE latent variables and (b) is a clustering result of the autoencoder latent variables.

The clustering results of the proposed method and the comparative method are shown in Figure 3. As can be seen in both results, the proposed AAE showed good clustering performances. However, in the autoencoder case, two defect classes were clustered into one cluster, with no clustering being done for one class. The adjusted mutual information score and homogeneity score for the two clustering results were 1.0 and 1.0, and 0.961 and 0.940, respectively, indicating better results in the AAE. These numerical results demonstrate that the proposed AAE model enhances the differences in the latent space, resulting in more accurate clustering.

4. Conclusion

In this study, we proposed a fully unsupervised clustering method based on AAE and Bayesian mixture model. We integrated the DP with AAE to estimate the prior distribution of AAE during learning. As a result, compared to general autoencoders, AAE was learned in a way that enhanced the differences between data points in the latent space. By performing unsupervised clustering on the mapped latent variables in a situation where the number of clusters was unknown, the Bayesian mixture model successfully achieved clustering. We believe that the proposed method will help overcome the problem of insufficient labels in real industries. However, this study was validated only for the CWRU dataset; additional experiments in other situations are needed. Furthermore, the relationship between the cluster index resulting from the proposed method and defect information is not studied, and further investigation is needed. Future research will focus on these points.

Acknowledgement

This work was supported in part by the Institute of Civil Military Technology Cooperation funded by the Defense Acquisition Program Administration and Ministry of Trade, Industry and Energy of the Korean government under grant No. 19-CM-GU-01, in part by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) Grant funded by the Korean Government [Ministry of Trade, Industry, and Energy (MOTIE)] under Grant 20206610100290, and in part by the K-CLOUD Research Project funded by the Korea Hydro & Nuclear Power Co. Ltd.

References

Hoang, D.-T. & Kang, H.-J. (2019). A survey on deep learning based bearing fault diagnosis. *Neurocomputing*, vol. 335, pp. 327-335.

Duan, Z. Wu, T. Guo, S. Shao, T. Malekian, R. & Li, Z. (2018). Development and trend of condition monitoring and fault diagnosis of multi-sensors information fusion

for rolling bearings: a review. *The International Journal of Advanced Manufacturing Technology*, vol. 96, no. 1, pp. 803-819.

Liu, R. Meng, G. Yang, B. Sun, C. & Chen, X. (2016). Dislocated time series convolutional neural architecture: An intelligent fault diagnosis approach for electric machine, *IEEE Transactions on Industrial Informatics*, vol. 13, no. 3, pp. 1310-1320.

Liu, H., Zhou, J., Zheng, Y., Jiang, W., & Zhang, Y. (2018). Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders. *ISA transactions*, 77, 167-178.

Makhzani, A., Shlens, J., Jaitly, N., Goodfellow, I., & Frey, B. (2015). Adversarial autoencoders. arXiv preprint arXiv:1511.05644.

Kim, T., & Lee, S. (2022). A Novel Unsupervised Clustering and Domain Adaptation Framework for Rotating Machinery Fault Diagnosis. IEEE Transactions on Industrial Informatics.

Case Western Reserve University Bearing Data Center, [Online]. Available: https://csegroups.case.edu/bearingdatacenter/home

Li, X., Jia, X. D., Zhang, W., Ma, H., Luo, Z., & Li, X. (2020). Intelligent cross-machine fault diagnosis approach with deep auto-encoder and domain adaptation. Neurocomputing, 383, 235-247.

Taewan Kim received the B.S. degree from Pohang University of Science and Technology, Pohang, South Korea, in 2020. He is currently pursuing the M.S./Ph.D. degree with the Industrial Artificial Intelligence Laboratory, Pohang University of Science and Technology, Pohang, South Korea. His research interests include industrial artificial intelligence with mechanical systems and applications of AI-based smart manufacturing.

Seungchul Lee received the B.S. degree in mechanical and aerospace engineering from Seoul National University, Seoul, South Korea; the M.S. and Ph.D. degrees in mechanical engineering from the University of Michigan, Ann Arbor, MI, USA, in 2008 and 2010, respectively. He has been an Associate Professor with the Department of Mechanical Engineering, Pohang University of Science and Technology, since 2018. His research focuses on Industrial Artificial Intelligence for Mechanical Systems, Smart Manufacturing, Materials, and Healthcare. He extends his research work to both knowledge-guided AI and AI-driven knowledge discovery.