Novel Graph-Based Features for Bearing Fault Diagnosis: Two Aspects of Time-Series Structure

Sangho Lee[1], Chihyeon Choi[2], and Youngdoo Son[3]

[1,2,3] *Dongguk University - Seoul, Seoul, Korea*
sangho218@dgu.ac.kr
choich0509@dgu.ac.kr
youngdoo@dongguk.edu

ABSTRACT

The feature-based methods for bearing fault diagnosis in prognostics and health management have been achieved satisfactory performances because of their robustness to noise and reduced dimension by pre-defined features. However, widely employed time- and frequency-domain features are insufficient to recognize the global pattern that indicates the structure of a time-series instance. In this paper, we propose two novel graph-based features which reflect the connection strength and degree of time series, respectively. First, we construct a graph of which an edge is defined as the Euclidean distance between each pair of time steps to measure the strengths of connections between the nodes. The other graph is constructed by the visibility algorithm, which converts a time series into a complex network to reflect the degrees of connections. Then, we calculate the Frobenius norms of the adjacency matrices of both graphs and use them as features for bearing fault diagnosis. To verify the proposed features, we performed several experiments with both synthetic and real datasets. From the synthetic datasets, it is observed that the changes in amplitudes and frequencies are detected by the features for the connection strength and degree, respectively. In addition, the proposed features also well-separate the distributions of each bearing state, including normal and several fault types, and show significant performance improvement as applied to the fault diagnosis task.

1. INTRODUCTION

As the complexity of equipment increases with industry development, the early detection of faults becomes important (Wei & Söffker, 2020). Feature-based methods for bearing fault diagnosis in prognostics and health management (PHM) have been achieved effective performances because of their robustness to noise and reduced dimension by pre-defined features (Ma, Zheng, Li, & Cottrell, 2019). However, traditional features require a lot of domain knowledge and are specialized in time-domain and frequency-domain. Although the traditional features can reflect the local relationship of the time series to some extent, it is difficult to reflect the global pattern of the time series (Ferreira & Zhao, 2016). A graph is a powerful mechanism to recognize the global pattern of a time series by identifying the relationship between data points or groups (Ferreira & Zhao, 2016; Aminikhanghahi & Cook, 2017), so graph-based methods have been introduced to reflect structural information.

T. Li et al. (2020) converted frequency information of time-series signal into an affinity graph and performed the gearbox fault diagnosis by applying a modified graph convolutional network. Zhou et al. (2021) introduced a framework for constructing graphs from time-series signals and using it to rotating machinery fault diagnosis. C. Li et al. (2020) constructed a graph by applying a horizontal visibility algorithm (Luque, Lacasa, Ballesteros, & Luque, 2009) to a time series and performed bearing fault diagnosis using a graph neural network. Wang et al. (2019) constructed a graph by deriving a frequency spectrum based on periodogram estimation for normal data, and utilized it to detect bearing fault using statistical analysis. However, the previous studies require sufficient training data and domain knowledge to diagnose the faults accurately. If they are insufficient, it is difficult to diagnose the fault correctly.

Recently, to solve these limitations, some research based on spectral graph theory has emerged. The spectral graph theory is the study to recognize the properties of a graph, such as characteristic polynomial, eigenvalues, and eigenvectors of adjacency matrices associated with the graph. These studies aim to detect faults under the assumption that there are structure changes of data between the normal and fault states of machinery. The eigenvalues and eigenvectors are used to detect two aspects of structural changes in the graph. First, a community structure change, namely connection strength, oc-

curs when strongly connected nodes are weakened or weakly connected nodes become strongly connected. On the other hand, a change in community activity, namely connection degree, happens when the number of connections between nodes increases or decreases. We can identify the changes in connection strength and degree by monitoring eigenvalues and eigenvectors, respectively (Wang, Lu, Liu, & Yan, 2018; Kannan, Vempala, & Vetta, 2004; Sarkar & Boyer, 1998). With this approach, the methods based on spectral graph theory show good performance in terms of early fault detection and generalization with insufficient data owing to the cycle-to-cycle strategy.

Sun et al. (2020) used the spectral graph theory as preprocessing for feature extraction. Specifically, they introduced a method that extracts fault features using maximum correlated kurtosis deconvolution. To improve the performance of fault feature extraction, they firstly constructed an adjacency matrix by calculating Euclidean distance between time steps and used graph similarity based on eigendecomposition to identify fault states in advance. Lu et al. (2018) constructed adjacency matrices for time-series instances in a normal state in the same way as in Sun et al. (2020), and derived representative eigenvector and eigenvalue using eigendecomposition for the averaged matrix. Then, with the fixed derived eigenvector, they used a martingale-test based on the Frobenius norm of the difference of non-diagonal component between the derived eigenvalue matrix and that of a time-series instance to be tested. However, these methods only reflect the connection strength without consideration the presence or absence of connection degree between time steps and have high complexity, $\mathcal{O}(n^3)$, where n denotes the number of time steps in a time series, due to the use eigendecomposition of the adjacency matrix. Thus, it is difficult to use in practice. In addition, when calculating the eigenvector for the average of adjacency matrices, there is a risk of information loss on the raw time series.

Therefore, we propose simple graph-based features that relieve the limitations above. Euclidean distance and visibility algorithm (Lacasa, Luque, Ballesteros, Luque, & Nuno, 2008) are applied to time series to construct two adjacency matrices that reflect the connection strength and degree of a time-series instance, respectively, and the norms of these matrices are used as graph-based features. These features have the following advantages:

- It helps to achieve good performance in terms of early fault detection and generalization even with insufficient data owing to the cycle-to-cycle strategy of the spectral graph theory.
- Information loss that may occur from the average of adjacency matrices (Lu et al., 2018) is eliminated by analyzing the features of each time-series instance separately.
- Detection delay, which is inevitable for the cycle-to-cycle strategy, is minimized by reducing the complexity of feature calculation.
- By reflecting not only the connection strength, but also the degree of connection between time steps, the structural information of the time series is sufficiently recognized.

2. PRELIMINARIES

In this section, we first explain traditional time- and frequency-domain features used to compare with the proposed features. Then, we briefly explain a visibility algorithm used to derive the connection degree of time series. Finally, Wasserstein and energy distances used to calculate the distance between class distributions are explained.

2.1. Time- and Frequency-Domain Features

Among the traditional time- and frequency-domain features presented in Jeon et al. (2015) and Jung et al. (2017), we used all (eight) time-domain features and chose three popular frequency-domain features that are the basis for calculating other frequency-domain features. The features used in this paper are presented in Table 1.

Table 1. The traditional time- and frequency-domain features used for fault diagnosis (Jeon et al., 2015; Jung et al., 2017). N is the number of time-series instances, t_n is a time-series instance, $\bar{t}$ is the sample mean of all time-series instances, and σ denotes the standard deviation. f and $S(\cdot)$ denote the frequency and power spectrum function, respectively.

Domain	Feature	Description
Time	Kurtosis	$\sum_{n=1}^{N} \frac{(t_n-\bar{t})^4}{\sigma^4}$
	Skewness	$\sum_{n=1}^{N} \frac{(t_n-\bar{t})^3}{\sigma^3}$
	Absolute Mean	$\sum_{n=1}^{N} \frac{\lvert t_n \rvert}{N}$
	Maximum	$max(\lvert t_n \rvert)$
	RMS	$\sqrt{\sum_{n=1}^{N} \frac{\lvert t_n \rvert^2}{N}}$
	Crest factor	$\frac{\text{Maximum}}{\text{RMS}}$
	Shape factor	$\frac{\text{RMS}}{\text{Absolute Mean}}$
	Impulse factor	$\frac{\text{Maximum}}{\text{Absolute Mean}}$
Frequency	FC	$\frac{\int f \times S(f) df}{\int S(f) df}$
	RMSF	$\sqrt{\frac{\int f^2 \times S(f) df}{\int S(f) df}}$
	RVF	$\sqrt{\frac{\int (f-\text{FC})^2 \times S(f) df}{\int S(f) df}}$

The time-domain features are obtained by using the raw vibration signal itself. First, there are two data statistics-related

features, kurtosis, and skewness. The kurtosis is a feature of the sharpness of the data distribution and is used to measure how intensively the instances are centered. The skewness represents the degree of asymmetry of data distribution. In addition, there are three commonly included kinetic energy-related features, maximum, absolute mean, and root mean square (RMS). The maximum is the maximum value of the signal, the absolute mean is the average of the absolute magnitude of the signal, and the RMS denotes the most suitable feature to quantify the magnitude of the signal. Finally, three sinusoidal wave shape-related features calculated using the above time-domain features are also included in time-domain features. On the other hand, we used three fundamental frequency-domain features for comparison with the proposed features. Frequency center (FC) and root mean square frequency (RMSF) are the scales indicating the change in the position of the fundamental frequency, and root variance frequency (RVF) indicates the degree of cohesion of the power spectrum.

2.2. Visibility Algorithm

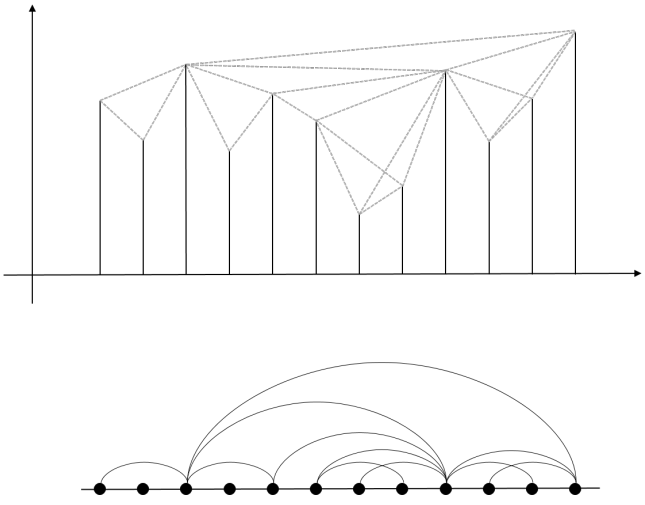


Figure 1. Overview of the visibility algorithm to obtain VG

Visibility algorithm converts a time series into a complex network (Lacasa et al., 2008). Figure 1 shows the overview of constructing visibility graph (VG) using the visibility algorithm. Several studies using VGs have proven that VGs help to effectively analyze time series and extract meaningful information (Y. Gao, Yu, & Wang, 2020; Z. Gao, Small, & Kurths, 2017). Let $T = \{t_1, t_2, \cdots, t_n\}$ be the set of time steps in a time series, and $Y = \{y_1, y_2, \cdots, y_n\}$ be the set of data values corresponding to time steps. Two arbitrary data points (t_a, y_a) and (t_b, y_b) are connected in a transformed graph, VG, if Eq. (1) is satisfied for all (t_c, y_c) placed between them.

$$y_c < y_b + (y_a - y_b) \times \frac{t_b - t_c}{t_b - t_a}, (a < c < b). \quad (1)$$

The resulting VG has the following properties (Lacasa et al., 2008):

- Connected: each adjacent node pair is connected
- Undirected: there is no directionality in the connection
- Invariant under affine transformations: the visibility criterion is invariant to affine transformations

In addition, we can recognize the time-series structure by analyzing the degree distribution of the resulting VG. For example, a periodic time series is converted into a regular graph, and a random time series are converted into an exponential random graph.

2.3. Distance Metrics for Distributions

In general, different classes (states) constitute separate manifolds. A feature that can well-distinguish each class distribution is a good feature that can provide useful information to a classifier (Bengio, Courville, & Vincent, 2013). To verify the usefulness of the proposed graph-based features, we measure the distances between class distributions derived by each feature. At this time, we use Wasserstein distance and energy distance to measure the distance between distributions (Arjovsky, Chintala, & Bottou, 2017; Shen, Qu, Zhang, & Yu, 2018; Bellemare et al., 2017).

The Wasserstein distance is a metric to measure the distance between two distributions and is defined as follows:

$$W(\mathbb{P}_X, \mathbb{P}_Y) = \inf_{\gamma \in \Pi(\mathbb{P}_X, \mathbb{P}_Y)} \mathbb{E}_{(x,y)\sim\gamma} \|x - y\|_1, \quad (2)$$

where $\Pi(\mathbb{P}_X, \mathbb{P}_Y)$ denotes the set of all possible joint distributions of which the marginal distributions are $\mathbb{P}_X$ and $\mathbb{P}_Y$. The Wasserstein distance between $\mathbb{P}_X$ and $\mathbb{P}_Y$ is defined as the minimum expected value of $\|x - y\|_1$ (Arjovsky et al., 2017).

The other metric, energy distance, is also used to measure the distance between two distributions. When two objects are located in the same positions in the gravitational space, the potential energy between the two objects is zero, and the potential energy increases as the distance between the two objects increases. The energy distance extends this concept to measure the distance between two distributions. When X and Y are independent random vectors with cumulative distribution functions F and G, respectively, the square of the energy distance between F and G is defined as Eq. (3) (Rizzo & Székely, 2016).

$$D^2(F, G) = 2\mathbb{E}\|X - Y\| - \mathbb{E}\|X - X'\| - \mathbb{E}\|Y - Y'\|, \quad (3)$$

where $\|\cdot\|$ is Euclidean norm for each component, X' and Y' are random variables identical to X and Y, respectively, which are independent and identically distributed. The energy distance between F and G is calculated by the square root of $D^2(F, G)$.

3. PROPOSED FEATURES

In this paper, we propose two novel graph-based features, *variability* and *connectivity*, for two different aspects of time-series structure. The procedures of calculating two graph-based features are summarized in Algorithm 1.

Algorithm 1 Novel graph-based features

Input: A time-series instance $T = \{t_1, \cdots, t_n\}$
Output: $\nu \in \mathbb{R}$: *variability*, $\kappa \in \mathbb{R}$: *connectivity*
An Euclidean-based adjacency matrix $G \in \mathbb{R}^{n \times n}$
A VG-based adjacency matrix $H \in \mathbb{R}^{n \times n}$
for $i, j = 1, \cdots, n$ **do**
Euclidean distance $d_{i,j} \leftarrow ||t_i - t_j||$
$G_{i,j} \leftarrow d_{i,j}$
Assign visibility $h_{i,j}$ by visibility criterion (Eq. (1)).
$H_{i,j} \leftarrow h_{i,j}$
end for
$\nu \leftarrow ||G||_F$, $\kappa \leftarrow ||H||_F$

First, we obtain two adjacency matrices representing two different aspects, strength and degree of connection, of the time-series structural information. Specifically, given a time-series instance $T = \{t_1, \cdots, t_n\}$, we derive an *Euclidean distance-based adjacency matrix* G corresponding to the time-series instance T, which is filled with the Euclidean distance values between all data point pairs ($||t_i - t_j||$). Thus, each element of G can represent the strength of the connection between data points. In addition, a VG-based adjacency matrix H corresponding to T is obtained. We apply the visibility criterion to each data point to construct H, so each element of H indicates whether t_i and t_j are connected.

Then, we introduce a formal definition of a graph-based feature, *variability*, that represents the strength of time-series connection.

Definition 1 (*Variability*) *Given an Euclidean distance-based adjacency matrix G corresponding to T, we define variability ν of T as*

$$\nu = ||G||_F, \tag{4}$$

where $||\cdot||_F$ denotes the Frobenius norm.

The *variability* feature, which is one aspect of the structural information of the time-series instance T, quantifies the overall strength that the data points in the time-series instance are connected. The Euclidean distance $d_{i,j}$, which is an element in the i-th row and j-th column of G, implies the connection strength between two data points t_i and t_j. Finally, the graph-based feature ν reflecting the overall connection strength of the time-series instance is extracted by calculating the Frobenius norm of G. The large *variability* value means that the Euclidean distances between all data points in the time-series instance are large; thus, we can conjecture that the time series has a large fluctuation.

On the other hand, we define a feature that reflects the degree of time-series connection as follows:

Definition 2 (*Connectivity*) *Given a VG-based adjacency matrix H corresponding to T, we define connectivity κ of T as*

$$\kappa = ||H||_F, \tag{5}$$

where $||\cdot||_F$ is the Frobenius norm.

The *connectivity* feature that represents another aspect of the time-series structural information identifies how many connections that satisfy the visibility criterion exists between the data points in the time-series instance T. An element in the i-th row and j-th column of H is assigned 1 if data points t_i and t_j have visibility. Thus, it can represent the degree of connection for the time-series instance. Similar to deriving the *variability* feature ν, we reflect the overall connection degree of T by calculating the Frobenius norm for H. The variation of *connectivity* value means that degree of connection is changed, so it means that the structure of the time series is changed.

The time complexity of the calculation of *variability* is $\mathcal{O}(n^2)$ because the Euclidean distances between all data points in a time-series instance should be computed, and that of the *connectivity* is also $\mathcal{O}(n^2)$ due to complexity of visibility algorithm. Therefore, we can consider the global pattern of the time series, which is difficult to be recognized by the traditional time- and frequency-domain features, using the proposed graph-based features with scalable complexities.

4. EXPERIMENTS

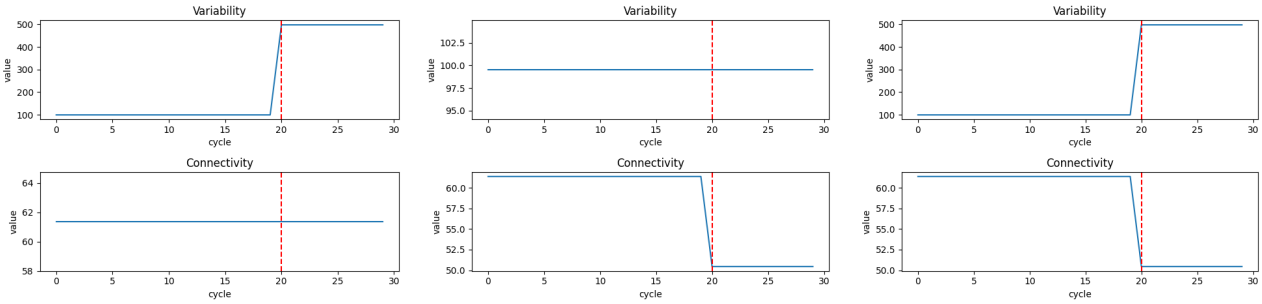
We performed three experiments to analyze the properties of the two proposed features, *variability* and *connectivity*, with a synthetic dataset and to verify their usefulness and applicability for bearing fault diagnosis with the real-world dataset. In each experiment, the proposed features were compared with the traditional time- and frequency-domain features.

4.1. Data Description

First, we constructed three synthetic time series to analyze the properties of the proposed features. Since amplitude and frequency are important properties in the vibration signal (Wang et al., 2018), three synthetic time series were constructed in which amplitude and frequency changes exist. Each data point t_i of each synthetic time series T is composed as follows:

$$t_i = \begin{cases} sin(2\pi i), & 1 \le i < 2000, \\ 5 \times sin(2\pi i), & 2000 \le i \le 3000. \end{cases}$$

$$t_i = \begin{cases} sin(2\pi i), & 1 \le i < 2000, \\ sin(4\pi i), & 2000 \le i \le 3000. \end{cases}$$



(a) Only amplitude change (b) Only frequency change (c) Amplitude and frequency change

Figure 2. The values of each proposed feature corresponding to each cycle when (a) only amplitude change, (b) only frequency change, and (c) amplitude and frequency change. The x-axis represents the value of each feature, and the y-axis is the cycle.

$$t_i = \begin{cases} sin(2\pi i), & 1 \leq i < 2000, \\ 5 \times sin(4\pi i), & 2000 \leq i \leq 3000. \end{cases}$$

The first and second synthetic time series reflected the only amplitude and frequency changes, respectively. In the last synthetic time series, the amplitude and frequency changes were reflected simultaneously. For the experiment in Section 4.2, each synthetic time series was divided into 30 by setting the sequence length of a time-series instance to 100.

Next, to verify the usefulness and applicability of the proposed features for bearing fault diagnosis, we used the Case Western Reserve University(CWRU) bearing dataset (Smith & Randall, 2015), widely used as a benchmark dataset for various studies to identify bearing states (Chen, Mauricio, Li, & Gryllias, 2020; X. Li, Zhang, & Ding, 2019; Zhang et al., 2020; Afrasiabi, Afrasiabi, Parang, & Mohammadi, 2019). The CWRU bearing dataset contains vibration signals representing the operation states from bearings. The vibration signals were collected at 12 or 48 kHz for bearings under four types of motor loads (0, 1, 2, or 3 hp). The corresponding rotating speed for each motor load is 1797, 1772, 1750, or 1730 rpm. In addition, there are four bearing states: 1) normal, 2) inner race, 3) outer race, and 4) ball faults. Each fault state has various diameters (0.007, 0.014, or 0.021 inches). For the experiments in Sections 4.3 and 4.4, we only used the vibration signals collected from the drive end at 48 kHz and used faults with 0.007 fault diameter. Moreover, the corresponding sequence length for each motor load was calculated to approximate one rotation cycle dividing sampling frequency by the rotating speed; hence, we set each sequence length to 1610, 1630, 1650, or 1670.

4.2. Results of Property Analysis

We performed change detection using the synthetic dataset to analyze the properties of the proposed features. Figure 2 shows the changes of *variability* and *connectivity* values overtime for three synthetic time series.

The proposed features, *variability* and *connectivity*, can detect amplitude and frequency changes, respectively. For the synthetic time series with only amplitude change, the value of the *variability* feature changes at the time step where the amplitude changes, whereas the value of the *connectivity* is maintained. Conversely, in the synthetic time series with only frequency change, only the value of the *connectivity* feature changes according to the change in frequency. When both amplitude and frequency change, we can observe that both the values of *variability* and *connectivity* change at the time of change.

4.3. Results of Usefulness Verification

A feature that can provide useful information should have a well-distinguished distribution for each class (Bengio et al., 2013). Therefore, for the CWRU dataset, we calculated the distance between the class distributions formed by each feature. At this time, min-max scaling was applied to each feature to remove the influence of the scales of features. We used the Wasserstein and energy distances as the distance metrics. In this experiment, we used data with 0 hp of motor load. The results of the Wasserstein and energy distances are shown in Table 2 and Table 3, respectively.

The *variability* has a property that can detect amplitude change of the vibration signal. In general, there is a significant difference in amplitude between fault types; hence, we can observe that the *variability* had the largest distance between fault class distributions (BI, BO, and IO). In addition, the max and RMS, which belong to the time-domain features that can reflect amplitude information, had larger distances between fault class distributions than that of the frequency-domain features. Conversely, there is a considerable difference in frequency between normal and fault classes (NB, NI, and NO). Thus, the *connectivity*, which can detect the frequency change of vibration signal, had a third largest distance between normal and fault class distributions. In this case, most frequency-domain features also had larger dis-

Table 2. Wasserstein distance between the class distributions derived by each feature. The three largest average distances are highlighted in boldface. (N, normal state; B, ball fault; I, inner race fault; O, outer race fault; F, fault states; AVG., average)

Domain	Feature	NB	NI	NO	NF AVG.	BI	BO	IO	FF AVG.
Time	Kurtosis	0.61	0.05	0.76	0.47	0.56	0.15	0.71	0.47
	Skewness	0.13	0.07	0.28	0.16	0.08	0.16	0.23	0.16
	Abs. Mean	0.43	0.07	0.88	0.46	0.35	0.45	0.80	0.54
	Max	0.41	0.04	0.86	0.44	0.36	0.46	0.82	**0.55**
	RMS	0.38	0.06	0.90	0.45	0.32	0.51	0.84	**0.56**
	Crest factor	0.51	0.07	0.42	0.33	0.44	0.09	0.35	0.30
	Shape factor	0.00	0.00	0.24	0.08	0.00	0.24	0.24	0.16
	Impulse factor	0.01	0.00	0.27	0.09	0.01	0.26	0.27	0.18
Frequency	FC	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	RMSF	0.71	0.82	0.61	**0.71**	0.11	0.10	0.21	0.14
	RVF	0.46	0.70	0.90	**0.69**	0.24	0.44	0.20	0.29
Graph	*Variability*	0.38	0.06	0.90	0.44	0.32	0.52	0.84	**0.56**
	Connectivity	0.53	0.53	0.83	**0.63**	0.01	0.30	0.30	0.20

Table 3. Energy distance between the distributions of each state derived by each feature. The three largest average distances are highlighted in boldface. (N, normal state; B, ball fault; I, inner race fault; O, outer race fault; F, fault states; AVG., average)

Domain	Feature	NB	NI	NO	NF AVG.	BI	BO	IO	FF AVG.
Time	Kurtosis	1.05	0.18	1.17	0.80	0.99	0.35	1.12	0.82
	Skewness	0.22	0.14	0.55	0.30	0.15	0.35	0.47	0.32
	Abs. Mean	0.91	0.37	1.30	0.86	0.82	0.91	1.24	0.99
	Max	0.88	0.27	1.29	0.81	0.82	0.90	1.25	**0.99**
	RMS	0.87	0.33	1.32	0.84	0.79	0.97	1.27	**1.01**
	Crest factor	0.89	0.19	0.83	0.64	0.82	0.22	0.75	0.60
	Shape factor	0.05	0.05	0.59	0.23	0.08	0.59	0.60	0.42
	Impulse factor	0.11	0.03	0.63	0.25	0.12	0.62	0.63	0.46
Frequency	FC	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	RMSF	1.14	1.23	1.06	**1.14**	0.33	0.38	0.60	0.44
	RVF	0.92	1.15	1.32	**1.13**	0.65	0.92	0.60	0.72
Graph	*Variability*	0.86	0.32	1.32	0.83	0.79	0.98	1.27	**1.01**
	Connectivity	0.53	0.53	0.83	**1.08**	0.01	0.30	0.30	0.48

Table 4. Accuracy (%) of the model trained various feature combinations of the time-, frequency-domain, and graph-based features. In each motor load, the best performance is highlighted in boldface.

Motor Load (hp)	Feature Combination						
	Time	Frequency	Graph	Time & Frequency	Time & Graph	Frequency & Graph	All
0	96.11	21.67	**99.72**	96.67	**99.72**	**99.72**	99.44
1	97.78	21.67	**100.00**	98.06	**100.00**	**100.00**	**100.00**
2	97.22	25.00	**97.22**	93.89	**97.22**	**97.22**	**97.22**
3	97.50	25.83	**100.00**	97.78	**100.00**	**100.00**	**100.00**

tances than the time-domain features. In summary, similar to the time- and frequency-domain features, the *variability* and *connectivity* can represent amplitude and frequency information of the vibration signal using two aspects of time-series structure, respectively. Moreover, we confirmed that the proposed features well-distinguished class distributions according to their properties demonstrated in Section 4.2.

4.4. Results of Applicability Verification

To verify that the proposed graph-based features, *variability* and *connectivity*, are adequate for bearing fault diagnosis, we

performed a classification task and compared the results of the model trained with various feature combinations, including the time-, frequency-domain, and graph-based proposed features. We used logistic regression as a classifier because it does not require additional parameter tuning. Table 4 shows the accuracy of each model, which was derived with a one-vs-rest strategy. In this experiment, we constructed a dataset with 30 instances randomly sampled per class, and 70% of the dataset was used to train the model and the rest to test model performance. To reduce the effect of randomness, we repeated the procedure ten times and reported the averaged results across all runs.

Although the number of the graph-based features is smaller than that of the other domains, the trained model only with the proposed features showed better performance than the trained model only with the traditional features, regardless of motor loads. Furthermore, when the traditional and proposed features were used together to train the model, the performance was improved compared to training the model only with the traditional features. It can be explained that the proposed features, *variability* and *connectivity*, play an important role in bearing fault diagnosis by reflecting structural information of time series, amplitude and frequency, while the traditional features only provide redundant information.

5. Conclusion

We propose novel graph-based features, *variability* and *connectivity*, for reflecting structural information of time series. We construct two graphs using the Euclidean distance and visibility algorithm and obtain the proposed features by calculating Frobenius norms of their adjacency matrices. Through several experiments on synthetic and real bearing datasets, we demonstrated that the *variability* and *connectivity* could reflect amplitude and frequency information, respectively, with reasonable complexities and well-separate the distributions of bearing states. The model trained only with the proposed features achieved significant performance in the bearing fault diagnosis task. Moreover, the proposed features helped improve the model performance trained with the other domain features. Therefore, the *variability* and *connectivity* features are useful features to classify bearing states in fault diagnosis.

Acknowledgment

This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT: Ministry of Science and ICT) (Nos. 2020R1C1C1003425 and 2020R1A4A3079710).

References

Afrasiabi, S., Afrasiabi, M., Parang, B., & Mohammadi, M. (2019). Real-time bearing fault diagnosis of induction motors with accelerated deep learning approach. In *2019 10th international power electronics, drive systems and technologies conference (pedstc)* (pp. 155–159).

Aminikhanghahi, S., & Cook, D. J. (2017). A survey of methods for time series change point detection. *Knowledge and information systems*, *51*(2), 339–367.

Arjovsky, M., Chintala, S., & Bottou, L. (2017). Wasserstein generative adversarial networks. In *International conference on machine learning* (pp. 214–223).

Bellemare, M. G., Danihelka, I., Dabney, W., Mohamed, S., Lakshminarayanan, B., Hoyer, S., & Munos, R. (2017). The cramer distance as a solution to biased wasserstein gradients. *arXiv preprint arXiv:1705.10743*.

Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, *35*(8), 1798–1828.

Chen, Z., Mauricio, A., Li, W., & Gryllias, K. (2020). A deep learning method for bearing fault diagnosis based on cyclic spectral coherence and convolutional neural networks. *Mechanical Systems and Signal Processing*, *140*, 106683.

Ferreira, L. N., & Zhao, L. (2016). Time series clustering via community detection in networks. *Information Sciences*, *326*, 227–242.

Gao, Y., Yu, D., & Wang, H. (2020). Fault diagnosis of rolling bearings using weighted horizontal visibility graph and graph fourier transform. *Measurement*, *149*, 107036.

Gao, Z., Small, M., & Kurths, J. (2017). Complex network analysis of time series. *EPL (Europhysics Letters)*, *116*(5), 50001.

Jeon, B. C., Jung, J. H., Youn, B. D., Kim, Y., & Bae, Y. (2015). Datum unit optimization for robustness of a journal bearing diagnosis system. *International Journal of Precision Engineering and Manufacturing*, *16*(11), 2411–2425.

Jung, J. H., Jeon, B. C., Youn, B. D., Kim, M., Kim, D., & Kim, Y. (2017). Omnidirectional regeneration (odr) of proximity sensor signals for robust diagnosis of journal bearing systems. *Mechanical Systems and Signal Processing*, *90*, 189–207.

Kannan, R., Vempala, S., & Vetta, A. (2004). On clusterings: Good, bad and spectral. *Journal of the ACM (JACM)*, *51*(3), 497–515.

Lacasa, L., Luque, B., Ballesteros, F., Luque, J., & Nuno, J. C. (2008). From time series to complex networks: The visibility graph. *Proceedings of the National Academy of Sciences*, *105*(13), 4972–4975.

Li, C., Mo, L., & Yan, R. (2020). Rolling bearing fault diagnosis based on horizontal visibility graph and graph neural networks. In *2020 international conference on sensing, measurement & data analytics in the era of*

artificial intelligence (icsmd) (pp. 275–279).

Li, T., Zhao, Z., Sun, C., Yan, R., & Chen, X. (2020). Multireceptive field graph convolutional networks for machine fault diagnosis. *IEEE Transactions on Industrial Electronics*, *68*(12), 12739–12749.

Li, X., Zhang, W., & Ding, Q. (2019). Understanding and improving deep learning-based rolling bearing fault diagnosis with attention mechanism. *Signal processing*, *161*, 136–154.

Lu, G., Liu, J., & Yan, P. (2018). Graph-based structural change detection for rotating machinery monitoring. *Mechanical Systems and Signal Processing*, *99*, 73–82.

Luque, B., Lacasa, L., Ballesteros, F., & Luque, J. (2009). Horizontal visibility graphs: Exact results for random time series. *Physical Review E*, *80*(4), 046103.

Ma, Q., Zheng, J., Li, S., & Cottrell, G. W. (2019). Learning representations for time series clustering. *Advances in neural information processing systems*, *32*, 3781–3791.

Rizzo, M. L., & Székely, G. (2016). Energy distance. *wiley interdisciplinary reviews: Computational statistics*, *8*(1), 27–38.

Sarkar, S., & Boyer, K. L. (1998). Quantitative measures of change based on feature organization: Eigenvalues and eigenvectors. *Computer vision and image understanding*, *71*(1), 110–136.

Shen, J., Qu, Y., Zhang, W., & Yu, Y. (2018). Wasserstein distance guided representation learning for domain adaptation. In *Thirty-second aaai conference on artificial intelligence.*

Smith, W. A., & Randall, R. B. (2015). Rolling element bearing diagnostics using the case western reserve university data: A benchmark study. *Mechanical systems and signal processing*, *64*, 100–131.

Sun, W., Zhou, Y., Cao, X., Chen, B., Feng, W., & Chen, L. (2020). A two-stage method for bearing fault detection using graph similarity evaluation. *Measurement*, *165*, 108138.

Wang, T., Lu, G., Liu, J., & Yan, P. (2018). Graph-based change detection for condition monitoring of rotating machines: Techniques for graph similarity. *IEEE Transactions on Reliability*, *68*(3), 1034–1049.

Wang, T., Lu, G., & Yan, P. (2019). A novel statistical time-frequency analysis for rotating machine condition monitoring. *IEEE Transactions on Industrial Electronics*, *67*(1), 531–541.

Wei, X., & Söffker, D. (2020). Comparison of cwru dataset-based diagnosis approaches: Review of best approaches and results. In *European workshop on structural health monitoring* (pp. 525–532).

Zhang, J., Yi, S., Liang, G., Hongli, G., Xin, H., & Hongliang, S. (2020). A new bearing fault diagnosis method based on modified convolutional neural networks. *Chinese Journal of Aeronautics*, *33*(2), 439–447.

Zhou, K., Yang, C., Liu, J., & Xu, Q. (2021). Dynamic graph-based feature learning with few edges considering noisy samples for rotating machinery fault diagnosis. *IEEE Transactions on Industrial Electronics*.

BIOGRAPHIES

Sangho Lee was born in Seoul, Republic of Korea in 1995. He received B.S. and M.S. degrees in Industrial and Systems Engineering from Dongguk University - Seoul, Seoul, Korea, in 2018 and 2020, respectively. He is currently pursuing the Ph.D. degree in Industrial and Systems Engineering at the Dongguk University-Seoul. His research interests include time-series analysis, complex network theory, machine learning, and their applications to PHM.

Chihyeon Choi was born in Busan, Korea in 1996. He received B.S. degrees in Industrial and Systems Engineering from Dongguk University - Seoul, Seoul, Republic of Korea, in 2022. He is currently pursuing the M.S. degree in Industrial and Systems Engineering at the Dongguk University-Seoul. His research interests include time-series analysis, learning with noisy labeled data, machine learning, and their applications to industrial problems.

Youngdoo Son is an assistant professor at the department of Industrial and Systems Engineering, Dongguk University - Seoul, Seoul, Korea. He received B.S. in Physics and M.S. in Industrial and Management Engineering from Pohang University of Science and Technology (POSTECH), Gyeongbuk, Republic of Korea in 2010 and 2012, respectively, and Ph.D. in Industrial Engineering from Seoul National University in 2015. His research interests include machine learning and data analytics, and their applications to industrial problems.