Research Article

Improvement of Roller Bearing Diagnosis with Unlabeled Data Using Cut Edge Weight Confidence Based Tritraining

Wei-Li Qin,[1] Wen-Jin Zhang,[1] and Zhen-Ya Wang[1,2]

[1]*School of Reliability and System Engineering, Beihang University, Beijing, China*
[2]*Science & Technology on Reliability & Environmental Engineering Laboratory, Beijing, China*

Correspondence should be addressed to Zhen-Ya Wang; wangzy@buaa.edu.cn

Roller bearings are one of the most commonly used components in rotational machines. The fault diagnosis of roller bearings thus plays an important role in ensuring the safe functioning of the mechanical systems. However, in most cases of bearing fault diagnosis, there are limited number of labeled data to achieve a proper fault diagnosis. Therefore, exploiting unlabeled data plus few labeled data, this paper proposed a roller bearing fault diagnosis method based on tritraining to improve roller bearing diagnosis performance. To overcome the noise brought by wrong labeling into the classifiers training process, the cut edge weight confidence is introduced into the diagnosis framework. Besides a small trick called suspect principle is adopted to avoid overfitting problem. The proposed method is validated in two independent roller bearing fault experiment vibrational signals that both include three types of faults: inner-ring fault, outer-ring fault, and rolling element fault. The results demonstrate the desirable diagnostic performance improvement by the proposed method in the extreme situation where there is only limited number of labeled data.

1. Introduction

Roller bearings are one of the most commonly used components in rotational machines and their faults may lead to huge economic losses, environment pollution, and human casualties. Hence, the fault diagnosis of the roller bearing is vital to guarantee the smooth and safe functioning of the mechanical systems.

There are a great deal of researches on vibration-based fault diagnosis of roller bearings and several powerful diagnostic methods are available [1]. Li et al. [2] presented an approach for motor roller bearing fault diagnosis using neural networks. Seryasat et al. [3] brought forward a ball bearing fault diagnosis method using fast Fourier transform (FFT) and wavelet energy entropy mean and root mean square (RMS). Peng and Chiang [4] used C4.5 decision tree and random forest algorithm to diagnose the fault of ball bearing of three-phase induction motor. Jin et al. [5] introduced a bearing fault diagnosis method using trace ratio linear discriminant analysis. And Liu et al. [6] proposed an extended wavelet spectrum analysis technique to achieve a more positive assessment of bearing health conditions. In fact, all these methods yield a rather excellent performance for fault diagnosis of different bearings. However, the data used in those methods are all labeled data, the kind that are already marked according to the bearing states. In the case of bearing fault diagnosis, however, the labeled data are quite expensive to obtain since they require human effort while large amount of unlabeled data is readily available. For a better practical value, the use of unlabeled data ought to be considered. Therefore, semisupervised learning, a technique that exploits unlabeled data plus few labeled data to train a good classifier, might be promising candidates in the area of roller bearing diagnosis when there is limited number of labeled data.

Good reviews [7, 8] have given out a good review of semisupervised classification methods. Among them, generative models, self-training, and cotraining are three classic semisupervised learning methods. Generative models specify a joint probability distribution over observation and label sequences and thus are used for modeling data. Nigam et al. applied the expectation maximization (EM) algorithm, a classic generative model, on mixture of multinomial distribution for the task of text classification. And the result showed the classifiers performed better than those trained only from

labeled data [9]. However, the generative model must be carefully constructed to reflect reality; otherwise unlabeled data that are supposed to help may actually hurt accuracy [10]. Self-training is a technique where a classifier is first trained from the small amount of labeled data and then used to classify the unlabeled data that will be added to the training set for further retraining. Rosenberg et al. [11] applied self-training to object detection systems from images and show the semisupervised technique compares favorably with a state-of-the-art detector. But self-training suffers from wrong labeling; note that the classifier uses its own predictions to teach itself [12]. Cotraining, proposed by Blum and Mitchell [13], can be quite effective, where in the extreme case only one labeled point is needed to learn the classifier, which is utmost incredibly amazing [14]. However, cotraining makes more than strong assumptions that (1) features can be split into two sets; (2) each subfeature set is sufficient to train a good classifier; and (3) the two sets are conditionally independent given the class on the splitting of features, which generally cannot be met in real life. To deal with this problem, Zhou and Li [15] proposed a cotraining style semisupervised learning algorithm named tritraining. In tritraining process, three weak classifiers are generated from the original labeled example set and are then refined using unlabeled examples. Tritraining neither requires the instance space to be described with sufficient and redundant views nor puts any constraints on the supervised learning algorithm. In addition, it possesses the merits of good efficiency and generalization ability. Tritraining has been successfully applied in Chinese chunking [16], biomedical named entity recognition [17], and web spam detection [18]. With all these advantages and successful application in other areas, tritraining is supposed to be a promising method in bearing fault diagnosis too. However, the process of unlabeled data adopted in tritraining is the simplistic consistency principle. In detail, in each round of tritraining, an unlabeled example is labeled for the third classifier if the other two classifiers agree on the labeling, under certain conditions. This might undermine the performance stability of tritraining because the unlabeled data may often be wrongly labeled by both classifiers during the learning process [19]. In order to overcome this problem, the cut edge weight statistic (CEWS) [20] is utilized to give the confidence of each predicted label of the unlabeled data. Only when the confidence is high enough can the predicted label be added to training set. With this problem solved by cut edge weight confidence (CEWC) plus all its merits, there is no doubt that tritraining will be a promising semisupervised algorithm for improvement of bearing fault diagnosis.

Hence to fully appreciate the large amount of unlabeled data of roller bearing and thus improve the performance of bearing fault diagnosis, this paper presents a roller bearing fault diagnosis method based on the combination of tritraining and CEWC. And the remainder of the paper is organized as follows. In Section 2, a detailed description of the methodologies used in this paper is presented. In Section 3, the experiment setup and relative information of two independent roller bearing fault datasets are presented. In Section 4, the results are presented. In Section 5, the results

are discussed. And finally in Section 6, the conclusion of the research is given.

2. Methodology

2.1. Tritraining. Tritraining is semisupervised machine learning proposed by Zhou and Li [15]. The procedure of tritraining is as follows. First three diverse classifiers are initially trained from the bagging samples from the original labeled example set. The diversity of the classifiers is guaranteed by the manipulation of the original labeled example set through a popular ensemble learning algorithm, that is, Bagging [21]. Second, the three trained classifiers are used to predict the examples from the unlabeled set. Those who pass the consistency principle will be added to the labeled dataset. Third, the initial classifiers are updated and the process repeats.

Let L denote the labeled dataset with size $|L|$ and U denote the unlabeled dataset with size $|U|$. In standard tritraining algorithm, there are three diverse classifiers h_1, h_2, and h_3 initially trained from the original L. Then, for any classifier, an unlabeled example can be labeled for it as long as the other two classifiers agree on the labeling of this example. For example, if h_1 and h_2 agree on the labeling of an example x in U, then x can be labeled for h_3. It is obvious that in such a scheme if the prediction of h_1 and h_2 on x is correct, then h_3 will get a solid new instance for further training. Otherwise, h_3 will get an example with noisy label. However, Zhou and Li [15] proved that, even in the worst case, the increase in the classification noise can be compensated if the amount of newly labeled examples is sufficient and the constraint condition (1) is met.

$$
\begin{aligned}
&\left| L \cup L^t \right| \left(1 - 2 \frac{\eta_L |L| + \hat{e}_1^t \left| L^t \right|}{\left| L \cup L^t \right|} \right)^2 \\
&> \left| L \cup L^{t-1} \right| \left(1 - 2 \frac{\eta_L |L| + \hat{e}_1^{t-1} \left| L^{t-1} \right|}{\left| L \cup L^{t-1} \right|} \right)^2,
\end{aligned}
\tag{1}
$$

where L^t and L^{t-1} are the set of examples that are labeled for a classifier by other two classifiers in the tth round and the $(t-1)$th round, respectively. $\hat{e}_1^t$ is the upper bound of the classification error rate of those other two classifiers in the tth round. And η_L is the classification noise rate of L; that is, the number of examples in L that are mislabeled is $\eta_L |L|$.

It is noteworthy that if the labeled examples are not sufficient or the constraint condition is not met, it is rather doubtable whether the benefits outweigh the drawbacks in case that an unlabeled example is wrongly labeled. Therefore, it is still necessary to measure the confidence of the labeling of each classifier.

2.2. Cut Edge Weight Confidence. The CEWC is established by a two-step process. In the first step, by employing the k-nearest neighbor criterion, a neighborhood graph is constructed from the labeled examples $L = \{(l_p, y_p) \mid p = 1, 2, \ldots, |L|\}$, where l_p is the attributes of pth example in set L and y_p the label. Concretely, each example $(l_p, y_p) \in L$
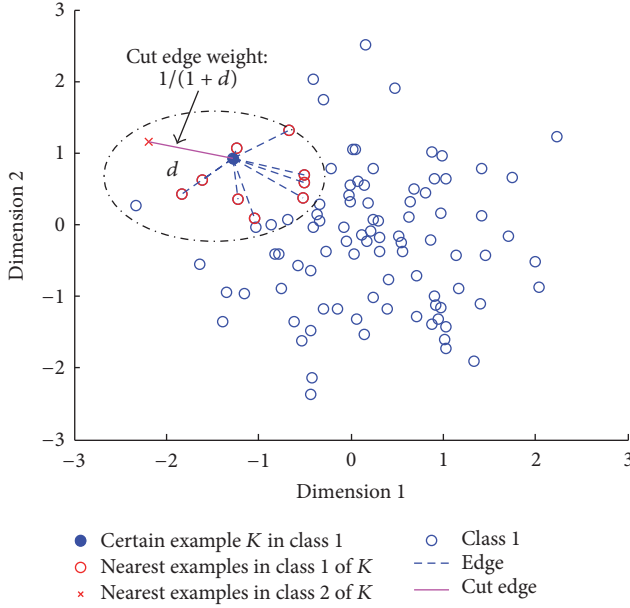
FIGURE 1: Cut edge illustration.

corresponds to a vertex in the graph G_L. There will be an edge $\overline{pq}$ connecting the two vertices of l_p and l_q if either l_p is among the k-nearest neighbors of l_q or l_q is among the k-nearest neighbors of l_p. And a weight $w_{pq} \in [0, 1]$ is associated with the edge $\overline{pq}$ computed as $(1 + d(l_p, l_q))^{-1}$, where $d(l_p, l_q)$ is the Euclidean distance between l_p and l_q.

In the second step, the confidence of whether the label y_p associated with l_p is correct is evaluated through exploring information encoded in G_L's structure. As illustrated in Figure 1, an edge in G_L is called *cut edge* if the two vertices connected by it have different associated labels. The CEWS is as follows:

$$J_p = \sum_{l_q \in C_p} w_{pq} I_{pq}, \tag{2}$$

where C_p corresponds to the set of examples which are connected with l_p in G_L and I_{pq} corresponds to an *i.i.d.* Bernoulli random variable which takes value of 1 if y_p is different from y_q. When the size of C_p is sufficiently large, according to the central limit theorem, J_p can be approximately modeled by a normal distribution. Let J_p^s denote the standardized form of J_p. Then based on the left unilateral p value of J_p^s with respect to $\mathbf{N}(0, 1)$, the labeling confidence of (l_p, y_p) is as follows:

$$CF_L(l_p, y_p) = 1 - \Phi(J_p^s), \tag{3}$$

where $CF_L(l_p, y_p)$ is the labeling confidence and $\Phi(J_p^s) = (1/\sqrt{2\pi}) \int_{-\infty}^{J_p^s} e^{-(t^2/2)} dt$ is the p value of J_p^s under standard normal distribution.

Note that $CF_L(l_p, y_p)$ represents only a heuristic way to estimate the labeling confidence of (l_p, y_p) and should by no means be deemed to represent the ground-truth probability

of y being the correct label of l. Though, experimental results in [22] validated the usefulness of this heuristic confidence estimation strategy in discriminating correctly labeled examples from incorrectly labeled ones.

2.3. Diagnosis Framework. The proposed approach combines the tritraining and CEWC to achieve bearing fault diagnosis and thus is called C-tritraining. The framework of it is illustrated in Figure 2. The data used for diagnosis are bearing vibration signals. First, the diagnostic features of the original vibration signals are extracted. Using ensemble empirical mode decomposition (EEMD) the original vibration signals can be broken down into intrinsic mode functions (IMFs) [23]. The information entropies of IMFs, which are surprisingly good features for bearing fault diagnosis [24], are used as the features, as the input of proposed method. Then, three bagging sample sets are drawn from the labeled feature set and each of them is used for the initial training of the weak classifier that we adopt BP neural network in this paper. Three weak classifiers will be obtained and used to predict certain proportion of unlabeled feature examples. In detail, the prediction of weak classifier 1 and weak classifier 2, if the CEWC of both are higher than the threshold, will be added to sample set 3 for updating of weak classifiers 3. The same goes for the updating of weak classifier 1 and classifier 2; that is, the training set is enlarged by the prediction of other two weak classifiers. Besides, the initial proportion of unlabeled features examples from the database is set to be 0.5. The proportion updates as follows:

$$\text{prop}_t = \text{prop}_{t-1} + (e_t - e_{t-1}), \tag{4}$$

where prop_t and prop_{t-1} are the proportion at tth and $(t-1)$th iteration. e_t and e_{t-1} are the training error at tth and $(t - 1)$th iteration. The proportion updating process is rather intuitive. If the error decreases after the enlargement of training set with added unlabeled prediction, we naturally are confident that the weak classifiers are reliable and able to handle more unlabeled examples next time. However, if the error increases, we will have lower confidence of the weak classifiers and less unlabeled examples are trusted to them next time. The tritraining process keeps running until the termination condition is reached. The final output of the framework is the ensemble classifier that will be used to do the final bearing diagnosis using majority voting.

Trying to avoid the overfitting problem, a small trick named suspect principle is introduced into the classifiers updating process as the termination condition. The core of suspect principle lies in that when the three initial classifiers in tritraining have been updated to their best (the error rates stop decreasing) with the help of unlabeled examples, we remain doubtful whether they have reached their best or just fall into a local optimum. Therefore, the termination condition is set as the classifiers updating process keeps running after certain times that the error rates stop decreasing. It is worth discussing how we should set the suspect principle value. The experimental results in Section 4 show that four times is a good choice.
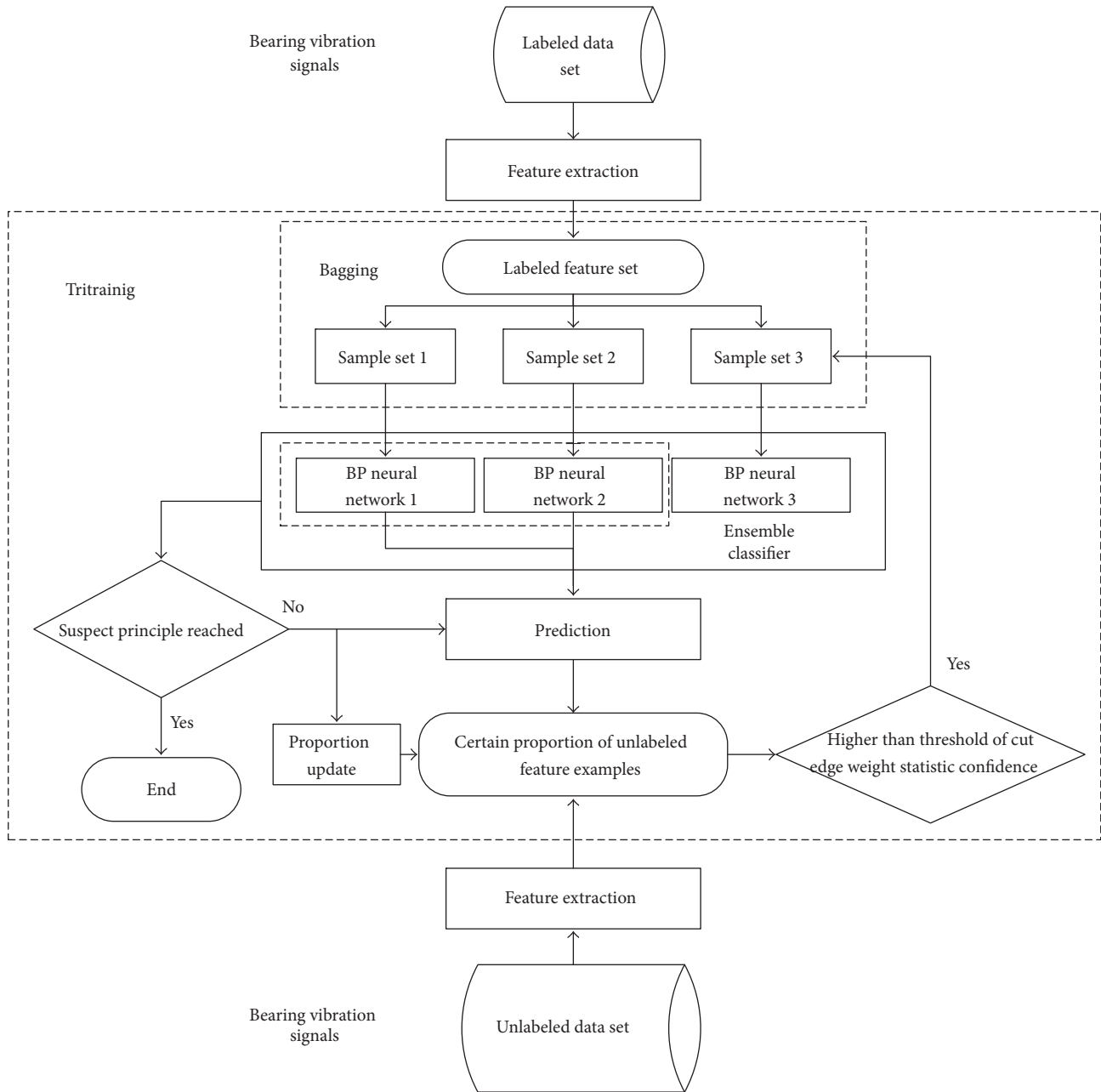
FIGURE 2: Framework of the proposed method.

3. Case Study Description

To verify the effectiveness and generalization ability of the proposed method, datasets from two individual bearing fault cases conducted by different groups were adopted.

Case 1. As shown in Figure 3, the first case was originally conducted on rotational machinery fault simulation test bed (QPZZ-II) by Prognostic and Health Management Laboratory of School of Reliability and Systems Engineering, BUAA.

The inner-ring fault, outer-ring fault, and roller element fault are introduced by wire-electrode cutting a crevice on the

surface of inner ring, outer ring, and one of the roller elements as marked in Figure 4. The vibrational signals are sampled at a frequency of 5120 samples per second and the rotation speed is 1500 revolutions per minute.

The test bearings used are cylindrical roller bearing (N205EM HRB CHINA), the detailed structure information of which is listed in Table 1.

Case 2. The second case was originally conducted by Institute of Intelligent instrument and Diagnosis, Xi'an Jiaotong University. The test rig shown in Figure 5 is completely designed and manufactured all by them. It consists mainly of a speed
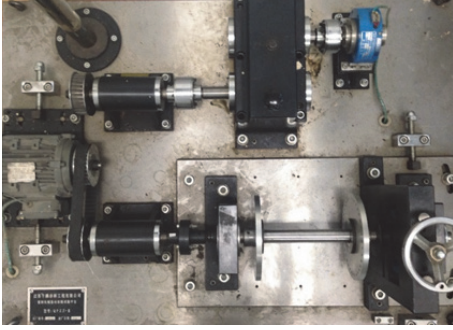
FIGURE 3: Test rig of Case 1.

TABLE 1: Bearing structure information.

Designation	Inside diameter	Outside diameter	Thickness	Element type
N205EM	25 mm	52 mm	15 mm	Cylindrical roller

TABLE 2: Bearing structure information.

Designation	Inside diameter	Outside diameter	Thickness	Element type
6308	40 mm	90 mm	23 mm	Deep groove ball

TABLE 3: Feature sets.

Feature set	Attribute	Size	Class	Nor./inner/outer/element
Case 1	24	400	4	25%/25%/25%/25%
Case 2	9	128	4	25%/25%/25%/25%

TABLE 4: Parameters of BPNN.

Number of layers	Input layer nodes	First hidden layer nodes	Second hidden layer nodes	Output layer nodes
4	24	8	4	4
Train function	*Function coefficient*	*Mean square error*	*Epoch*	*Validation checks*
Levenberg-Marquardt	$1.00e + 10$	$1.00e - 20$	1000	6

governor, driving motor, power supply box, horizontal and radial loading devices, and of course sensors.

Bearing faults in Case 2 include inner-ring fault, outer-ring fault, and roller element fault with an area of 3.8 mm^2, 7 mm^2, and 3 mm^2 circle-shaped spalling on the surface of inner ring, outer ring, and roller element, respectively. The test bearings used are deep groove ball bearing (6308), the detailed structure information of which is listed in Table 2. The sampling frequency is 20 K samples per second and the rotation speed is 1500 revolutions per minute.

4. Results

Through EEMD process, the original vibration signals collected from the two cases are transformed into two feature sets. According to [18], two parameters of EEMD, that is, the ratio of the standard deviation of the added noise and that of input, are set to be 0.15 and ensemble number is set to be 100. Information of the feature sets is tabulated in Table 3. For each feature set, 85 percent of the data are kept as training set while the rest are used as the test set to examine the trained classifiers. The training set, composed of labeled pool and unlabeled pool, that is, $L \cup U$, is partitioned under different *unlabeling rates* including 80 percent, 60 percent, 40 percent, and 20 percent. Take the data of Case 1 whose size is 400 examples; for example, the training set has 340 examples (85 percent) and test set 60 examples (15 percent). When the unlabeling rate is 80 percent, 68 examples out of

340 examples are then put into L and other 292 examples are put into U without their labels. To overcome the randomness of the results, 50 independent runs are performed and the averaged results are summarized as the final outcome.

Figure 6 shows the classification error rate of Cases 1 and 2 under different unlabeling rate and suspect principle value. When suspect principle value is set to four, the classification error rates are the lowest or the second lowest in most situations except only for the classification error rate in Case 2 under the unlabeling rate of 0.6. Therefore, it is naturally intuitive to determine that suspect principle value set to four is a practical optimal choice.

With suspect principle value set to be four, the averaged results are summarized in Table 4, which presents the classification error rate of the initial ensemble weak classifiers, that is, the combination of the three initial BP neural network classifiers only trained from L and the final ensemble classifiers generated by tritraining and the improvement of the latter over the former. The architecture and parameters of the BP neural network are shown in Table 4.

4.1. Comparative Experiments with Different Semisupervised Learning Models. In this paper, self-learning and tritraining models were conducted for comparison. The self-learning model is a traditional semisupervised learning method where the most confident unlabeled data samples, together with the predicted labels, are added to the initial training set, so that the neural network classifier can be retrained and the procedure repeats. The tritraining model is an elementary model whose parameters were the same as those of the C-tritraining except the CEWS optimization process. Detailed diagnosis is listed in Table 5–7 and Figure 7.

4.2. Comparative Experiments with Different Base Classifiers. For the purpose of investigating the diagnosis performance with different base classifiers, an additional experiment was conducted where the support vector machine (SVM) was built with a RBF kernel function whose kernel parameter is set to 0.08 and the penalty factor set to 128. The SVM model was trained using the one-versus-all criterion. Note that the SVM model could be regarded as a more stable classifier while

FIGURE 4: Faults of inner ring, outer ring, and roller element.

TABLE 5: Classification error rate of initial and final hypothesis and the corresponding improvements of C-tritraining under different unlabeling rate with suspect principle value set to 4.

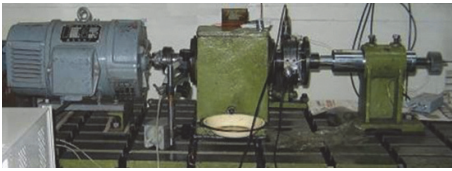| Case	Unlabeling rate of 0.8			Unlabeling rate of 0.6			Unlabeling rate of 0.4			Unlabeling rate of 0.2		
	Initial	Final	Improvement	Initial	Final	Improvement	Initial	Final	Improvement	Initial	Final	Improvement
Case 1	0.1670	0.1367	18.14%	0.0990	0.0857	13.43%	0.0613	0.06	2.12%	0.0500	0.0487	2.6%
Case 2	0.4589	0.3400	25.91%	0.3274	0.28	14.48%	0.2495	0.2211	11.38%	0.2453	0.2200	10.31%

FIGURE 5: Test rig of Case 2.

neural network based classifiers are mostly unstable in terms of training mechanism. Taking Case 1 as an example, detailed diagnosis results are displayed in Table 8.

5. Discussion

(1) Different from the supervised learning based diagnosis methods for fault detection and identification, this manuscript proposes a new incremental learning approach that takes advantage of unlabeled data to improve diagnosis performance of rolling bearings. Considering that fault samples are continuously attained over monitoring time, semisupervised ensemble learning is employed so as to avoid manual labeling error, as well as improving classification accuracy for health assessment utilizing prior learned knowledge and newly attained information in a real-time diagnosis mechanism. In this regard, tritraining, where three diverse classifiers are generated from the bagging samples and integrated for fault diagnosis, is conducted to improve classification performance of base classifiers. On this basis, CEWC is employed in this study to further mine salient characteristics of unlabeled data with a view to design a more intelligent diagnosis model. This method was applied

to two bearings with different proportions of unlabeled samples (20, 40, 60, and 80 percent, resp.). As shown in Table 5, the proposed method is able to effectively improve the performance of the initial ensemble classifiers under all unlabeled rates for both Cases 1 and 2. The improvement percentage ranges from 25.9% to 2.6%.

(2) It is noteworthy in Figure 8 that the improvement percentage increases sharply as unlabeling rate increases in both Cases 1 and 2. This means, by utilizing unlabeled data, the proposed method really makes a difference where there is limited labeled data to train the classifiers. And when there is less labeled data to train the classifiers, the proposed method is able to more improve the classifiers' performance. However, the absolute value of improvement and diagnostic error rate in Case 2 are commonly higher than those in Case 1. The difference between two results is caused by their dataset size. The feature set of Case 1 has 400 examples while Case 2 has only 128 examples. For Case 2, when unlabeling rate is 0.8, then there are only $128 \times 85\% \times 20\% \approx 22$ labeled examples to train classifiers, which is apparently not enough to train good classifiers. No wonder the initial underfitting classifiers' classification error reaches 0.4589 when unlabeling rate is 0.8. The proposed method promotes 25.91% performance of the initial classifiers in this extreme situation. When there is enough labeled data, for example, Case 1 when unlabeling rate is 0.2, the classification error rate lowers to 0.0487 (95.13% diagnostic accuracy). It implies that roller bearing fault diagnosis based on tritraining is promising in either situation (a) where there is not enough labeled data to obtain good classifiers or situation (b) where
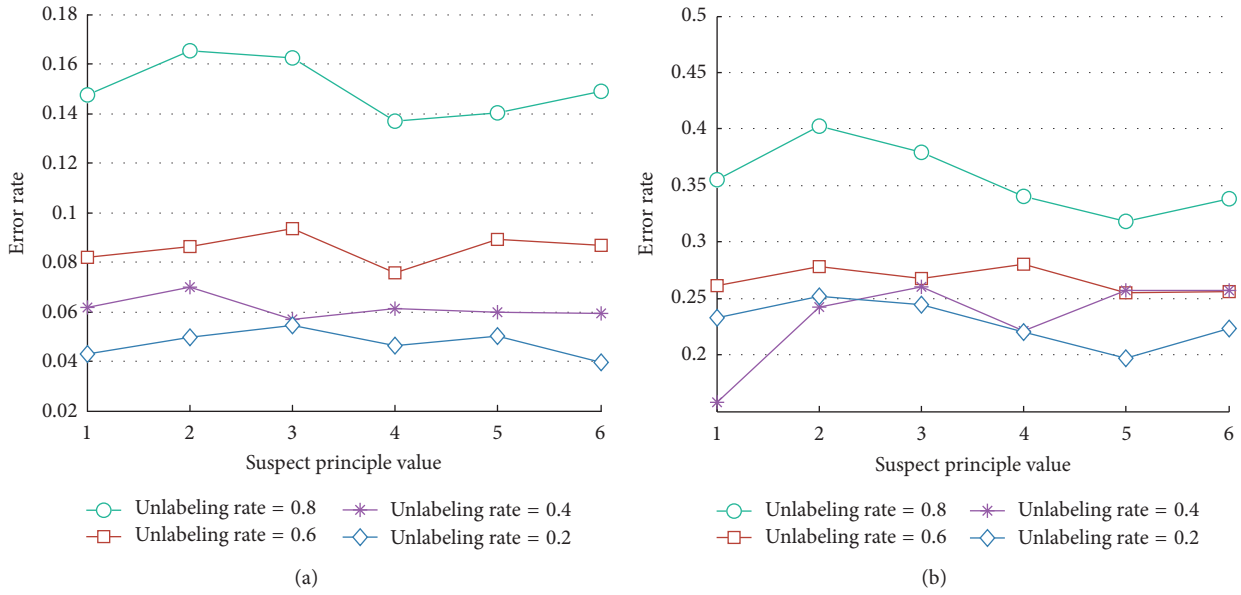
FIGURE 6: Classification error rate of Case 1 (a) and Case 2 (b) under different unlabeling rate and suspect principle value.
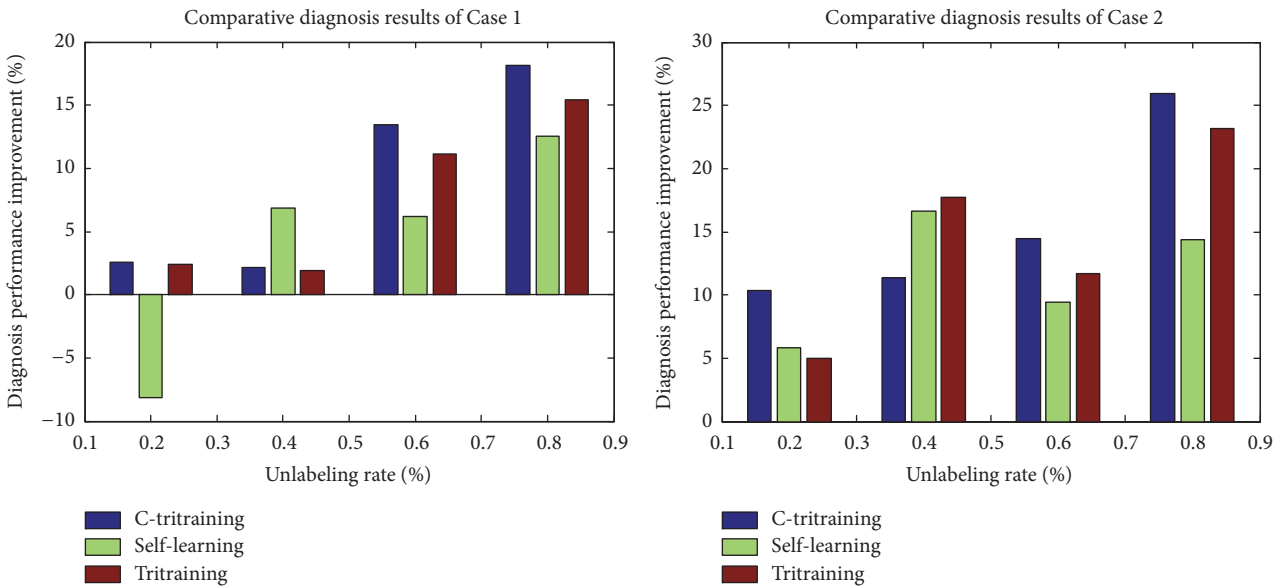

FIGURE 7: Performance improvement at different unlabeling rate in Cases 1 and 2.

there is enough labeled data. In situation (a) tri-training greatly improves classifiers' performance by utilizing unlabeled data that are easily available. And the performance will keep upgrading as long as there is more unlabeled data. In situation (b) tritraining can still be helpful though the initial classifiers are good enough for the bearing fault diagnosis.

(3) In this study, the semisupervised learning methods including self-learning and traditional tritraining are conducted for comparison. Detailed diagnosis results are listed in Tables 3–7 and Figure 8. It is observed that although the classification rates of all methods were improved, the tritraining based methods appeared to

produce higher correct rates in most of the cases. Taking the diagnosis results with unlabeling rate of 0.8 in Case 2 as an example, the improvements of classification accuracies were 25.91%, 14.33%, and 23.13% for C-tritraining, self-learning, and tritraining, respectively. This is mainly because such ensemble process could effectively strengthen the learning ability through integrating multiple views of individual classifiers. In addition, compared to basic tritraining model, the proposed method attained better diagnosis results demonstrating the effectiveness of CEWS on capturing pivotal fault characteristics from unlabeling data in rolling bearing diagnosis issues. It is also noted

TABLE 6: Classification error rate of initial and final hypothesis and the corresponding improvements of self-training under different unlabeling rate.

Case	Unlabeling rate of 0.8			Unlabeling rate of 0.6			Unlabeling rate of 0.4			Unlabeling rate of 0.2		
	Initial	Final	Improvement	Initial	Final	Improvement	Initial	Final	Improvement	Initial	Final	Improvement
Case 1	0.3353	0.2933	12.53%	0.2168	0.2033	6.23%	0.1002	0.0933	6.89%	0.0417	0.0454	−8.15%
Case 2	0.4737	0.4058	14.33%	0.4037	0.3658	9.39%	0.3158	0.2632	16.66%	0.2811	0.2648	5.80%

TABLE 7: Classification error rate of initial and final hypothesis and the corresponding improvements of tritraining under different unlabeling rate.

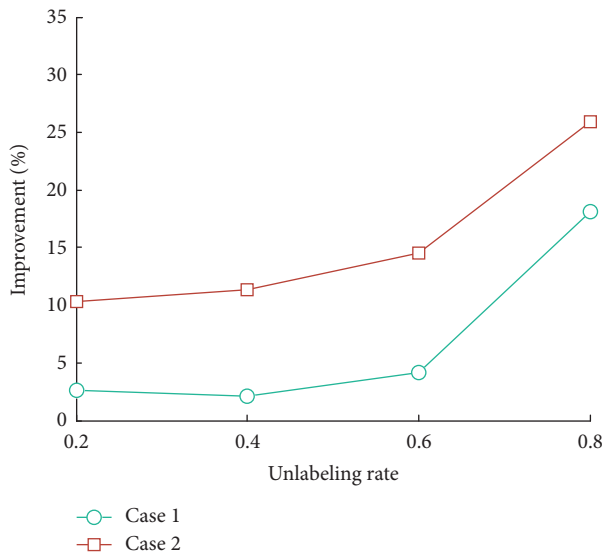| Case	Unlabeling rate of 0.8			Unlabeling rate of 0.6			Unlabeling rate of 0.4			Unlabeling rate of 0.2		
	Initial	Final	Improvement	Initial	Final	Improvement	Initial	Final	Improvement	Initial	Final	Improvement
Case 1	0.2167	0.1833	15.41%	0.1500	0.1333	11.13%	0.0667	0.0654	1.95%	0.0533	0.0520	2.44%
Case 2	0.3684	0.2832	23.13%	0.3632	0.3205	11.73%	0.2559	0.2105	17.74%	0.2105	0.2000	4.99%

FIGURE 8: Comparative diagnosis results of rolling bearing diagnosis cases.

in Case 1 that the diagnosis result of self-learning decreased with unlabeling rate of 0.8, which may due to some negative effects of improper training such as overfitting problems.

(4) From the diagnosis results of SVM based C-tritraining, it is noted that the fault classification performance was improved as well, demonstrating the effectiveness of the proposed semisupervised learning method in rolling bearing diagnosis; that is, such model can be appropriately applied using various base classifiers. However, the misclassification rates of the testing data were relatively high when compared to the BPNN based model, which may be due to the lesser difference between three SVM models. The ensemble process could only be more effective on condition that the base classifiers are of greater diversity. Therefore in this study, when determining the base classifier and architecture, we follow a simple

idea that the classifiers should be as different as possible in the bagging process so that more sufficient information could be learned from the unlabeling data.

6. Conclusion

In order to improve performance of bearing fault diagnosis when there is limited labeled data, this paper presents a roller bearing fault diagnosis method based on the combination of tritraining and CEWC. The method is validated in two roller bearing fault cases conducted by two independent groups. The results showed that, with the help of unlabeled examples, the method is able to effectively improve the fault diagnosis for both cylindrical roller bearing and deep groove ball bearing when there is limited labeled examples. The proposed method still helps even when there is enough labeled data and the diagnostic accuracy can reach up to 95%.

Although the proposed method is promising, there is something that could be improved in the future work. The feature extracted from the vibrational signal is information entropy of IMFs through EEMD, which is an iterative process and so is tritraining. That makes the proposed method time-consuming, which undermines its applicability in roller bearing online diagnosis. Hence, the efficiency improvement is among the priorities in future work.

Competing Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Authors' Contributions

Wei-Li Qin and Zheng-Ya Wang contributed equally to this work and should be considered joint first authors.
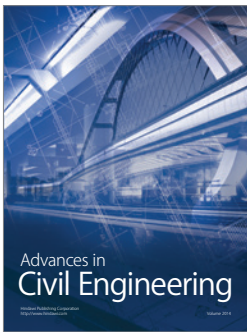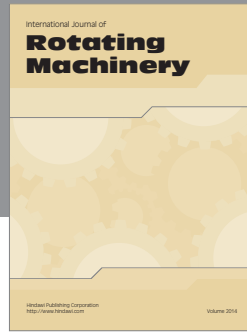
Acknowledgments

Table 8: Comparative diagnosis results of SVM and BPNN in Case 1.

Classifier	Unlabeling rate of 0.8			Unlabeling rate of 0.6			Unlabeling rate of 0.4			Unlabeling rate of 0.2		
	Initial	Final	Improvement	Initial	Final	Improvement	Initial	Final	Improvement	Initial	Final	Improvement
BPNN	0.1670	0.1367	18.14%	0.0790	0.0757	4.18%	0.0613	0.0600	2.12%	0.0500	0.0487	2.60%
SVM	0.3001	0.2617	12.80%	0.1697	0.1652	2.65%	0.1267	0.1250	1.34%	0.0833	0.0783	6.00%

References

[1] R. B. Randall and J. Antoni, "Rolling element bearing diagnostics—a tutorial," *Mechanical Systems & Signal Processing*, vol. 25, no. 2, pp. 485–520, 2011.

[2] B. Li, M.-Y. Chow, Y. Tipsuwan, and J. C. Hung, "Neural-network-based motor rolling bearing fault diagnosis," *IEEE Transactions on Industrial Electronics*, vol. 47, no. 5, pp. 1060–1069, 2000.

[3] O. R. Seryasat, M. A. Shoorehdeli, F. Honarvar, and A. Rahmani, "Multi-fault diagnosis of ball bearing using FFT, wavelet energy entropy mean and root mean square (RMS)," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC '10)*, pp. 4295–4299, Istanbul, Turkey, October 2010.

[4] H. W. Peng and P. J. Chiang, "Control of mechatronics systems: ball bearing fault diagnosis using machine learning techniques," in *Proceedings of the 8th Asian Control Conference*, pp. 175–180, May 2011.

[5] X. Jin, M. Zhao, T. W. S. Chow, and M. Pecht, "Motor bearing fault diagnosis using trace ratio linear discriminant analysis," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 5, pp. 2441–2451, 2014.

[6] J. Liu, W. Wang, and F. Golnaraghi, "An extended wavelet spectrum for bearing fault diagnostics," *IEEE Transactions on Instrumentation & Measurement*, vol. 57, no. 12, pp. 2801–2812, 2008.

[7] X. Zhu, "Semi-supervised learning literature survey," *Computer Science*, vol. 37, pp. 63–77, 2008.

[8] N. N. Pise and P. Kulkarni, "A survey of semi-supervised learning methods," in *Proceedings of the International Conference on Computational Intelligence and Security (CIS '08)*, vol. 2, pp. 30–34, Suzhou, China, December 2008.

[9] K. Nigam, A. K. Mccallum, S. Thrun, and T. Mitchell, "Text classification from labeled and unlabeled documents using EM," *Machine Learning*, vol. 39, no. 2, pp. 103–134, 2000.

[10] F. G. Cozman, I. Cohen, and M. C. Cirelo, "Semi-supervised learning of mixture models," in *Proceedings of the International Conference on on Machine Learning*, pp. 41–65, 2003.

[11] C. Rosenberg, M. Hebert, and H. Schneiderman, "Semi-supervised self-training of object detection models," in *Proceedings of the 7th IEEE Workshop on Applications of Computer Vision/IEEE Workshop on Motion and Video Computing (WACV/MOTION '05)*, pp. 29–36, Breckenridge, Colo, USA, January 2005.

[12] O. Chapelle, B. Schölkopf, and A. Zien, *Semi-Supervised Learning*, MIT Press, 2010.

[13] A. Blum and T. Mitchell, "Combining labeled and unlabeled data with co-training," in *Proceedings of the 11th Annual Conference on Computational Learning Theory*, pp. 92–100, 2000.

[14] O. Chapelle, B. Scholkopf, and A. Zien, "An augmented PAC model for semi-supervised learning," *Semi-Supervised Learning*, pp. 397–419, 2006.

[15] Z.-H. Zhou and M. Li, "Tri-training: exploiting unlabeled data using three classifiers," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 11, pp. 1529–1541, 2005.

[16] W. Chen, Y. Zhang, and H. Isahara, "Chinese chunking with tri-training learning," in *Computer Processing of Oriental Languages. Beyond the Orient: The Research Challenges Ahead: 21st International Conference, ICCPOL 2006, Singapore, December 17–19, 2006. Proceedings*, vol. 4285 of *Lecture Notes in Computer Science*, pp. 466–473, Springer, Berlin, Germany, 2006.

[17] Y. H. Cai and X. Cheng, "Biomedical named entity recognition with tri-training learning," in *Proceedings of the 2nd International Conference on Biomedical Engineering and Informatics (BMEI '09)*, pp. 1–5, Tianjin, China, October 2009.

[18] H. Li, "Web spam detection based on improved tri-training," in *Proceedings of the 2nd IEEE International Conference on Progress in Informatics and Computing (PIC '14)*, pp. 61–65, IEEE, Shanghai, China, May 2014.

[19] A. Blum and S. Chawla, "Learning from labeled and unlabeled data using graph mincuts," in *Proceedings of the 18th International Conference on Machine Learning*, pp. 19–26, 2001.

[20] F. Muhlenbach, S. Lallich, and D. A. Zighed, "Identifying and handling mislabelled instances," *Journal of Intelligent Information Systems*, vol. 22, no. 1, pp. 89–109, 2004.

[21] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.

[22] M.-L. Zhang and Z.-H. Zhou, "CoTrade: confident co-training with data editing," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 41, no. 6, pp. 1612–1626, 2011.

[23] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Advances in Adaptive Data Analysis*, vol. 1, no. 1, pp. 1–41, 2009.

[24] W.-L. Qin, W.-J. Zhang, and C. Lu, "Rolling bearing fault diagnosis based on ensemble empirical mode decomposition, information entropy and random forests," in *Proceedings of the International Conference on Vibroengineering*, pp. 211–216, September 2015.