Neurocomputing 168 (2015) 119–127

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Multimodal deep support vector classification with homologous features and its application to gearbox fault diagnosis

Chuan Li [a,b,*], René-Vinicio Sanchez [b], Grover Zurita [b], Mariela Cerrada [b], Diego Cabrera [b], Rafael E. Vásquez [c]

[a] Research Center of System Health Maintenance, Chongqing Technology and Business University, Chongqing 400067, China
[b] Department of Mechanical Engineering, Universidad Politécnica Salesiana, Cuenca, Ecuador
[c] Department of Mechanical Engineering, Universidad Pontificia Bolivariana, Medellín, Colombia

A R T I C L E I N F O

Article history:
Received 15 February 2015
Received in revised form
11 April 2015
Accepted 2 June 2015
Communicated by Shen Yin
Available online 16 June 2015

Keywords:
Deep learning
Support vector classification
Multimodal homologous feature
Gearbox
Fault diagnosis

A B S T R A C T

Gearboxes are crucial transmission components in mechanical systems. Fault diagnosis is an important tool to maintain gearboxes in healthy conditions. It is challenging to recognize fault existences and, if any, failure patterns in such transmission elements due to their complicated configurations. This paper addresses a multimodal deep support vector classification (MDSVC) approach, which employs separation–fusion based deep learning in order to perform fault diagnosis tasks for gearboxes. Considering that different modalities can be made to describe same object, multimodal homologous features of the gearbox vibration measurements are first separated in time, frequency and wavelet modalities, respectively. A Gaussian-Bernoulli deep Boltzmann machine (GDBM) without final output is subsequently suggested to learn pattern representations for features in each modality. A support vector classifier is finally applied to fuse GDBMs in different modalities towards the construction of the MDSVC model. With the present model, "deep" representations from "wide" modalities improve fault diagnosis capabilities. Fault diagnosis experiments were carried out to evaluate the proposed method on both spur and helical gearboxes. The proposed model achieves the best fault classification rate in experiments when compared to representative deep and shallow learning methods. Results indicate that the proposed separation–fusion based deep learning strategy is effective for the gearbox fault diagnosis.

1. Introduction

Gearboxes are key components in mechanical transmission systems. A gearbox fault may result in unwanted fatal breakdowns, production losses and even human casualties [1]. Hence any gearbox's defective components such as gears, bearings, and shafts should be diagnosed as early as possible [2]. Due to complicated configuration of the gearbox, it is challenging to recognize fault existences and, if any, the failure patterns.

Three basic steps are commonly used to diagnose gearbox fault patterns. The first one is to choose condition symptoms for the gearbox to be diagnosed. Such condition symptoms as vibration, acoustic, oil debris, electrical current, and heat measurements have been used for the gearbox fault diagnosis [3–5]. Within these symptoms, vibration analysis is the most commonly-used technique for condition monitoring and performance monitoring [6,7]. Although the vibration signal can be combined with other signals for failure recognition [8], in most cases only vibration measurements are used due to their cheap expense and simple implementation.

As condition symptoms are often confused by noises and interferences, raw signal of vibration measurements are usually insensitive to component defects, especially to incipient faults which are present at the early stage [9]. There, a second step to determine failure-sensitive features is usually needed and applied. In vibration based fault diagnosis, the most commonly-used features have been generated from temporal, spectral, wavelet [10–12], and other signal representations. Such different representations can be regarded as different observations of the vibration signal, and from a specific point of view each observation is a modality of the vibration signal [13]. Although they come from the same source, i.e., the vibration signal, different modalities may lead to different sensitivities for failure patterns [14]. In the time modality, Raad et al. [15] employed cyclostationarity as an indicator for gear diagnosis. In the frequency modality, Li and Liang [16] suggested an optimal mathematical morphology demodulation technique to extract impulsive feature for bearing defect diagnosis. In the wavelet modality, different statistical parameters have been introduced to classify fault patterns [17]. Features with different modalities have been reported effective for the fault

* Corresponding author. Tel.: +86 23 6276 8469.
E-mail address: chuanli@21cn.com (C. Li).

http://dx.doi.org/10.1016/j.neucom.2015.06.008
0925-2312/

diagnosis, but none of the modality can outperform others in all cases. Hence, some researchers suggested to observe the fault information in wider fashion, i.e., with multiple modalities. For instance, two diagnostic parameters from both time and frequency modalities have been reported for planetary gearbox diagnosis based on the examination of vibration characteristics [18].

With the extraction of defect-sensitive features by using one or more modalities, it is still not easy to specify a condition pattern due to the similarity between different fault patterns. For real applications, engineers may employ the diagnosed fault pattern information to prepare special measures for gearbox maintenances. Hence, building an effective classifier is always the third required step used by the fault diagnosis community. Tayarani-Bathaie et al. [19] developed a dynamic neural network as a classifier for gas turbine fault diagnosis. Guo et al. [20] suggested using a support vector machine in collaboration with envelope spectrum analysis to classify three health conditions of planetary gearboxes. For the gearbox fault diagnosis, Chen et al. [21] proposed an intelligent diagnosis model based on wavelet support vector machine and immune genetic algorithm. Among all the typical classifiers, the support vector classification (SVC) family (i.e., the standard SVC and its variants) attracted much attention due to its extraordinary classification performance. Recently, different data-driven approaches have been used for industrial processes [22,23]. The most popular and traditional classifiers are belong to “shallow” learning category [24]. Comparing with conventional methods, deep learning received great success in the classification due to its “deeper” representations for faulty features. Up to now, different deep learning networks such as deep belief network [25], deep Boltzmann machines (DBMs) [26], deep autoencoders [27], and convolutional neural networks [28] have been reported; however, few have been used for fault diagnosis cases. Tran et al. [29] introduced the application of the deep belief networks to diagnose reciprocating compressor valves. Tamilselvan and Wang [30] employed the deep belief learning based health state classification for iris dataset, wine dataset, Wisconsin breast cancer diagnosis dataset, and Escherichia coli dataset. Limited reports show deep learning structures for the fault diagnosis using one-modality features.

The deep learning represents an object in a “deep” style, while multiple modalities can observe the object in a “wide” fashion. This philosophy inspires the authors to propose here a multimodal deep support vector classification (MDSVC) with homologous features for the gearbox fault diagnosis. From the real world point of view, the fault diagnosis has been applied for different industrial cases. For the gearboxes, it is helpful for detecting the early defects and monitoring the healthy conditions [31,32]. The proposed method is capable of improving diagnosis accuracy from limited vibration signal sources. Although with greater computational burden during model training by the deep learning, application procedure using the trained model is time-saving. The proposed method employs a separation–fusion based deep learning strategy. Due to limited availability of vibration signals, one vibration measurement of a gearbox can be represented in time, frequency and wavelet modalities, separately. In each modality, a Gaussian-Bernoulli deep Boltzmann machine (GDBM), without final output, is suggested to learn the pattern information from homologous features. To integrate “wide” modalities with “deep” learning, a SVC is applied to fuse the GDBMs in different modalities as the MDSVC model. The proposed model is evaluated in gearbox fault diagnosis experiments and is compared with other state-of-the-art shallow and deep learning methods.

The rest of this paper is organized as follows. The MDSVC approach and its elements are introduced in Section 2. Section 3 presents in detail the application of the present MDSVC model to the gearbox diagnosis. In Section 4, fault diagnosis experiments based on two gearbox set-ups are carried out to evaluate the proposed method. Finally, some conclusions are addressed in Section 5.

2. Multimodal deep learning with support vector classification

In this section, deep learning with the GDBM is first introduced. For better accommodating multiple modalities of homologous features, a multimodal structure is subsequently proposed. To combine outputs of a multimodal structure, a data fusion technique based on the SVR is suggested for developing the MDSVC model.

2.1. Deep learning with GDBM classifier

The deep learning with different networks have claimed state-of-the-art performances in different tasks. The deep learning tries to hierarchically learn deep features from input data with very deep networks. The deep network is usually layer-wise initialized with unsupervised training, followed by supervised fine-tuning for the whole model. In this way, more abstract and complex features can be extracted at higher layers. Appropriate features can be therefore formulated for the classification at the top of the model.

Recent researches indicate that deep learning models are capable of producing better approximation for nonlinear functions comparing to “shallow” models. The DBM is a prominent type of the deep learning model. For a DBM, each layer represents complicated correlations between hidden features in the layer below. Hence the DBM has the potential of learning internal representations that become increasingly complex, highly desirable for solving different classification tasks. In this paper, therefore, the DBM is chosen as the basic network for our MDSVC model.

The standard DBM is a network of symmetrically couple stochastic binary neurons. As shown in Fig. 1, a single visible layer $\mathbf{v}$ and L hidden layers $\mathbf{h}^{(1)}$,…, $\mathbf{h}^{(l)}$, …, and $\mathbf{h}^{(L)}$ contribute a DBM network, where the connections are only allowed between the visible neurons and the first hidden ones, as well as between hidden neurons in adjacent hidden layers. The energy E of the state $\{\mathbf{v}, \mathbf{h}^{(1)}, \dots, \mathbf{h}^{(L)}\}$ is defined as [33]

$$E(\mathbf{v}, \mathbf{h}^{(1)}, \dots, \mathbf{h}^{(L)} \mid \theta) = -\sum_{i=1}^{N_v}\sum_{j=1}^{N_1} W_{ij} v_i h_j^{(1)} - \sum_{i=1}^{N_v} b_i v_i - \sum_{i=1}^{N_v}\sum_{j=1}^{N_l} b_j^{(l)} h_j^{(l)} - \sum_{i=1}^{L-1}\sum_{j=1}^{N_l}\sum_{k=1}^{N_l+1} w_{jk}^{(l)} h_j^{(l)} h_k^{(l+1)}, \quad (1)$$

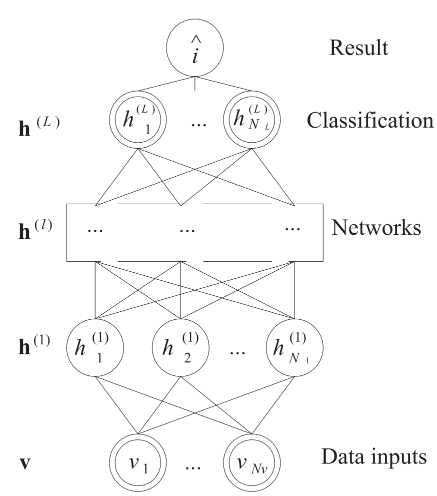


Fig. 1. Structure of the GDBM classifier.

where $\theta = \{\mathbf{W}, \mathbf{b}\}$ are model parameters, W_{ij} denotes the weight of the synaptic connection between the ith visible neuron and the jth hidden neuron, b_i represents the ith bias term, N_v is the number of visible neurons, and N_l stands for the number of neurons in the lth hidden layer.

The inputs for the standard DBM are limited to binary neurons. For modeling real-valued cases, a popular technique is to normalize the real-valued data into [0, 1] with a probability. This method, fits best to bounded variables but fails in dealing with unbounded ones. For the image processing where pixels are bounded between [0, 255], this strategy works well. However, it is very difficult to define lower and upper limits for fault diagnosis features. To tackle this issue, Cho et al. [34] proposed a Gaussian-Bernoulli deep Boltzmann machine (GDBM) which used Gaussian neurons in the visible layer of the DBM. Taking the standard deviations σ of visible neurons into consideration, Eq. (1) can be adopted for the GDBM as

$$E(\mathbf{v}, \mathbf{h}^{(1)}, ..., \mathbf{h}^{(L)} | \theta) = -\sum_{i=1}^{N_v}\sum_{j=1}^{N_1} W_{ij} v_i h_j^{(1)} / \sigma_i^2 + \sum_{i=1}^{N_v} 2(v_i - b_i)^2 / \sigma_i^2$$
$$-\sum_{i=1}^{N_v}\sum_{j=1}^{N_l} b_j^{(l)} h_j^{(l)} - \sum_{i=1}^{L-1}\sum_{j=1}^{N_l}\sum_{k=1}^{N_l+1} w_{jk}^{(l)} h_j^{(l)} h_k^{(l+1)} \cdot \quad (2)$$

The joint distribution over the visible and hidden layers is given by

$$p\left(\mathbf{v}, \left\{\mathbf{h}^{(l)}\right\} \middle| \theta\right) = \frac{1}{Z(\theta)} \exp\left(-E(\mathbf{v}, \left\{\mathbf{h}^{(l)}\right\} \middle| \theta)\right), \quad (3)$$

where $Z(\theta)$ denotes the normalizing constant dependent on the model parameters θ of the network.

With the modified energy function as given by Eq. (2), the conditional probability of the ith visible neuron can be derived as

$$p(v_i | \mathbf{h}^{(1)}, \theta) = N\left(v_i | \sum_{j=1}^{N_1} h_j^{(1)} W_{ij} + b_i, \sigma_i^2\right), \quad (4)$$

where $N(.|\mu, \sigma^2)$ is a probability density of normal distribution, with mean μ and standard deviation σ. Similarly, the conditional probabilities of neurons in different hidden layers are formulated as

$$p(h_j^{(l)} | \mathbf{h}^{(l-1)}, \mathbf{h}^{(l+1)}, \theta) = s\left(\sum_{i=1}^{N_{l-1}} h_i^{(l-1)} W_{ij}^{(l-1)} + \sum_{k=1}^{N_{l+1}} h_k^{(l+1)} W_{jk}^{(l)} + b_j^{(l)}\right), \quad (5)$$

where $s(.)$ is a sigmoid function. It is noted that there are two special cases, i.e., the last and the first hidden layers for the above equation. For the last hidden layer (i.e., $l = L$), we set $N_{L+1} = 0$. As for the first hidden layer (i.e., $l = 1$), parameters for Eq. (5) should be set as

$$\mathbf{h}^{(l-1)} = \mathbf{v}, \quad \text{and} \quad \sum_{i=1}^{N_{l-1}} h_i^{(l-1)} W_{ij}^{(l-1)} = \sum_{i=1}^{N_v} v_i W_{ij} / \sigma_i^2, \quad \text{for} \quad l = 1 \cdot \quad (6)$$

For the application of the GDBM specified for classification problems, it is common to use the softmax or 1-of-K encoding at the top. The total input $\mathbf{a}$ into the softmax layer can be specified as

$$a_i = \sum_{j=1}^{N_{L-1}} h_j^{(L-1)} W_{ij}^{(L-1)}, \quad (7)$$

To classify the given number (i.e., N_L) of the possible classes, the softmax layer should have c nodes with discrete probabilities, for simplicity, $p_1, \ldots, p_i, \ldots$, and p_{N_L} as given by

$$p_i = \frac{\exp(a_i)}{\sum_{i=1}^{N_L} \exp(a_i)}, \quad \text{and} \quad \sum_{i=1}^{N_L} p_i = 1 \cdot \quad (8)$$

The classified class $\hat{i}$ would accordingly be

$$\hat{i} = \arg\max_i p_i = \arg\max_i a_i \quad (9)$$

2.2. GDBM classification with multimodal homologous features

In the real world, the information to depict an object comes from multiple input channels, which form multimodal homologous features for the object. For example, both images and texts can be used to be linked with a meaningful identity. On the other hand, different words are applied to describe a same picture. Those different words, i.e., different modalities, might result in the same picture. In machinery fault diagnosis, a vibration measurement can be expressed in time, frequency and time–frequency domains. Those representations (or, modalities) in different domains should correspond to the same condition pattern. It is possible that one modality is insufficient to reflect all aspects of the object, and multimodal homologous data are helpful to discover a lot of features about the object.

Srivastava and Salakhutdinov [35] proposed a multimodal learning method with the DBMs. For a two-modality case, an additional layer is added on the top of the two DBMs. Let's suppose that the number of the hidden layers for the first DBM is L_1, and that for the second DBM is L_2. The conditional probabilities of the neurons in the top layer L are given by

$$p(h_j^{(L)} | \mathbf{h}^{(L_1)}, \mathbf{h}^{(L_2)}, \theta) = s\left(\sum_{i=1}^{N_{L1}} h_i^{(L1)} W_{ij}^{(L1)} + \sum_{k=1}^{N_{L21}} h_k^{(L2)} W_{jk}^{(L2)} + b_j^{(L)}\right) \cdot \quad (10)$$

And the classified class $\hat{i}$ can be also extracted using Eq. (9) where p_i should be replaced by $p(h_j^{(L)})$ given by Eq. (10).

Comparing Eq. (10) with Eq. (5) reveals that the modal learning suggested in Ref. [35] is a direct combination of multiple modalities. Nevertheless, the linear combination may introduce much computational burden for the learning. In this paper, we propose a more simple multimodal learning strategy for the GDBMs dealing with classification problems.

As shown in Fig. 2, the multimodal learning structure proposed in this paper can be regarded as an extension of Ref. [35]. Instead of a direct combination in the additive layer L, we suggest using a data fusion strategy for the multimodal learning. With the

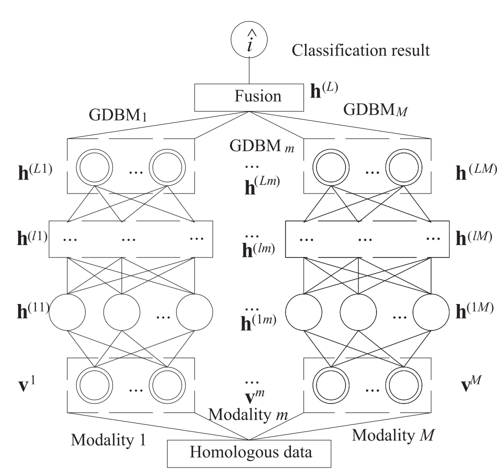


Fig. 2. Structure of the GDBM classifier with multimodal homologous features.

proposed structure, the learning procedure will be also simplifier than the direct combination.

With the proposed multimodal networks as shown in Fig. 2, one can revisit Eqs. (9) and (10) to obtain the classified class as

$$\hat{i} = F(p_1^{(L1)}, \dots, p_{N_{L1}}^{(L1)}, \dots, p_1^{(Lm)}, \dots, p_{N_{LM}}^{(LM)}) \tag{11}$$

where $F(.)$ is the fusion function to be used. One may notice that one of the crucial layers is the fusion layer $\mathbf{h}^{(L)}$. In the classically-used multimodal learning $\mathbf{h}^{(L)}$ is, actually, a linear combination layer, which is insufficient for dealing with nonlinear, complicated data fusion. For this reason, we propose in this paper to employ the support vector classifier as a tool for data fusion in multimodal learning, which will be detailed in the next section.

2.3. Support vector fusion for enhancing the multimodal GRBMs

A support vector machine (SVM) is a statistical learning tool that uses structural risk minimization theory to solve multi-dimensional problems. Yin et al. [36] proposed a robust 1-class SVM for the fault detection. The simulation shows that the addressed robust 1-class SVM is superior to the general one. The classical SVM can be formulated as a binary classifier, i.e., the SVC. The purpose of the support vector fusion is to employ the SVC for replacing the function $F(.)$ given by Eq. (11). To this end, the training data for the SVC should be

$$\mathbf{x}_n = (p_1^{(L1)}, \dots, p_{N_{L1}}^{(L1)}, \dots, p_1^{(Lm)}, \dots, p_{N_{LM}}^{(LM)})_n \quad \text{and} \quad y_n = (\hat{i})_n, \tag{12}$$

where n denotes the number of the observed data, $\mathbf{x}_n$ the input data, and y_n the classification output. Supposing the classification for y_n is binary, the standard SVC training is constrained by an optimization problem expressed as [37]

$$\begin{aligned} &\min_{\mathbf{w},\xi_n} \tfrac{1}{2}\mathbf{w}^{\mathbf{T}}\mathbf{w} + C\sum_{n=1}^{N}\xi_n \\ &\text{s.t.} \quad \mathbf{w}^{\mathbf{T}}\mathbf{x}_n\tau_n \geq 1-\xi_n \\ &\qquad\; \xi_n \geq 0, \quad n=1,2,\dots N, \end{aligned} \tag{13}$$

where $\tau_n \in \{-1,1\}$, ξ_n is the nth positive slack variable, C is a penalty parameter which controls a tradeoff between margin maximization and error minimization. The above constrained problem can be changed as a unconstrained one with a L2 norm expressed as follows:

$$\min_{\mathbf{w}} \frac{1}{2}\mathbf{w}^{\mathbf{T}}\mathbf{w} + C\sum_{n=1}^{N} \max\,(1-\mathbf{w}^{\mathbf{T}}\mathbf{x}_n\tau_n, 0)^2 \cdot \tag{14}$$

By solving the above equation, the predicted class can be obtained with

$$\hat{i} = \arg\max_{\tau}(\mathbf{w}^{\mathbf{T}}\mathbf{x})\tau, \tag{15}$$

Based on the aforementioned illustration, the standard SVM can be employed as a binary classifier. The fault diagnosis, however, often deals with the classification between multiple patterns. To tackle this problem, a one-against-all (OAA) strategy can be used for i class faulty patterns. In the OAA strategy, i SVMs are trained independently, where data from other classes form negative cases. Let the output of the ith SVM be

$$a_i(\mathbf{x}) = \mathbf{w}^{\mathbf{T}}\mathbf{x} \cdot \tag{16}$$

The predicted class from multiple optional patterns can be therefore recast as

$$\hat{i} = \arg\max_{i} a_i(\mathbf{x}) \cdot \tag{17}$$

It should be pointed out that Tang [38] developed a deep leaning algorithm using linear SVM approach and won facial expression recognition contest in the International Conference on Machine Learning 2013 REPL workshop. Our work is inspired by, but slightly differs from, his research. On the one hand, Tang's method employed penultimate activation (i.e., $\mathbf{h}^{(L-1)}$) as the input $\mathbf{x}$. In the present MDSVC model, the support vector classifier is used as an additional layer $\mathbf{h}^{(L)}$. On the other hand, Tang's method employed the SVM to replace the softmax regression, while the present MDSVC model introduce the SVM for the data fusion of the multiple softmax regression results. Nevertheless, the classification performances of the both models can be benefited from the SVM.

3. Application to gearbox fault diagnosis

After the development of the MDSVC model, in this section we employ the MDSVC as an effective tool for the gearbox fault diagnosis. Multimodal homologous features are first generated from vibration measurements. The present MDSVC approach is then applied to facilitate gearbox fault diagnosis using multimodal homologous features.

3.1. Multimodal homologous features of vibration measurements

Vibration signals of gearbox can be expressed in different modalities such as time, frequency, and the wavelet modalities. As introduced before, observations in different modalities represent different fashions of the gearbox vibration measurement. More importantly, different modalities come all from one source, i.e., the vibration measurement. Hence, multimodal homologous features can be generated from different modalities of the gearbox vibration measurement.

Letting $s(t)$ denote the vibration signal, its Fourier spectrum is given by

$$S(f) = \int_{-\infty}^{+\infty} s(t)e^{-2\pi jft}dt, \tag{18}$$

where t is time and f represents frequency. As in the wavelet modality, there are different wavelet transforms available. Considering its enhanced signal decomposition capability in high frequency region, and the comparatively low dimensions of decomposition numbers, wavelet packet transform (WPT) is used in this work as the wavelet modality presentation.

With an integral scale parameter u and translation parameter q ($q=0, \dots, 2^u-1$; $u=0, \dots, U$), a wavelet packet (WP) function $P_{u,q}^r(t)$ is given by [39]

$$P_{u,q}^r(t) = 2^{u/2}P^r(2^u t - q), \tag{19}$$

where $r=0, 1, \dots$ is oscillation parameter. The vibration signal $s(t)$ can be decomposed into U levels. At the uth level, there are 2^u packets with the order $r=1, 2, \dots, 2^u$. The WP coefficients $e_{u,q}^r$ in a WP node are inner product between the signal and the WP functions, i.e.,

$$e_{u,q}^r = < s(t), P_{u,q}^r > = \int_{-\infty}^{\infty} s(t)P_{u,q}^r(t)d(t), \tag{20}$$

where $\langle *,* \rangle$ denotes the inner product operator. In this way, for simplicity, we index a WP node as (u, r) with coefficients given by $e_{u,q}^r$. To represent the WPT of the vibration signal in a matrix fashion, one can flat WP nodes at all the levels and expressed as

$$WT(b, s(t)) = [\underbrace{e_{1,q}^1, e_{1,q}^2, \dots, e_{u,q}^{2^u}}_{b=1,2,\dots,2^{U+1}-1}]; \quad b \in R^{2^{U+1}-1}, \tag{21}$$

where n_0 represents the length of $s(t)$, and b stands for total number of the WP nodes.

Since representations in the three modalities are all very long, much computational burden will be encountered upon accommodating them using deep learning. Statistical variables have been approved simple yet effective features for the fault diagnosis community. Hence we employ the statistical variables as the observed features of the vibration signal in different modalities. In this work, such statistical variables are expressed as [40]

$$A_1(z) = \frac{\max |z(i)|}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(z(i))^2}}, \tag{22}$$

$$A_2(z) = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(z(i))^2}}{\frac{1}{N}\sum_{i=1}^{N}|z(i)|}, \tag{23}$$

$$A_3(z) = \frac{1}{N}\sum_{i=1}^{N}|z(i)|, \tag{24}$$

$$A_4(z) = \left(\frac{1}{N}\sum_{i=1}^{N}\sqrt{|z(i)|}\right)^2, \tag{25}$$

$$A_5(z) = \frac{1}{N}\sum_{i=1}^{N}(z(i))^4, \tag{26}$$

$$A_6(z) = \frac{1}{N}(z(i))^2, \tag{27}$$

$$A_7(z) = \frac{\max |z(i)|}{\left(\frac{1}{N}\sum_{i=1}^{N}\sqrt{|z(i)|}\right)^2}, \tag{28}$$

$$A_8(z) = \frac{\max |z(i)|}{\frac{1}{N}\sum_{i=1}^{N}|z(i)|}, \tag{29}$$

$$A_9(z) = \frac{1}{N}(z(i))^3 \cdot \tag{30}$$

where N is the length of the representation $z(i)$, and $A_1, A_2, \ldots, A_9$ stand for the crest factor, shape factor, absolute mean amplitude, square root amplitude, kurtosis, variance value, clearance factor, impulse indicator, and skewness factor of $z(i)$, respectively, and $z(i)$ represents an uniform of different representations, i.e.,

$$z(i) = s(t), S(f), WP(1, s(t)), \ldots, WP(2^{U+1}-1, s(t)) \cdot \tag{31}$$

In this way, multimodal homologous features include 9 time modality features, 9 frequency modality ones, and $9\times(2^{U+1}-1)$ wavelet modality ones. The features in the three modalities can be formulated by

$$z_1(i) = s(t),\ z_2(i) = S(f) \quad \text{and}$$
$$z_3(i) = \{WP(1, s(t)), \ldots, WP(2^{U+1}-1, s(t))\} \tag{32}$$

3.2. Application of the proposed MDSVC approach

After representing the vibration measurement in the three modalities, the MDSVC model can be developed using the structure shown in Fig. 2. With the developed MDSVC model, one should train the classification networks for the gearbox fault diagnosis. The MDSVC model training is divided into three stages: pretraining for each constituting GDBM, fine-tuning for each constituting GDBM, and training for the SVC.

For each constituting GDBM, a greedy, layer-by-layer unsupervised learning algorithm is applicable [41]. As neurons in intermediate layers of a GDBM receive information both from upper and lower layers, special attention should be paid for the pretraining period. To deal with this problem, Salakhutdinov [42] proposed a strategy to halve the pretrained weights in intermediate layers and to duplicate the visible and topmost layers.

After the unsupervised pretraining for all the GDBMs, for better fitting the fault diagnosis task, all the weights $\mathbf{W}$ should be discriminatively fine-tuned. Back-propagation (BP) is a commonly-used fine-tuning algorithm [43]. During the fine-tuning stage, training labels (i.e., $\hat{i}$) are directly used to train each GDBM structure. This is in the same way as for the fine-tuning of the standard GDBM.

With the pretraining and the fine-tuning processes, each GDBM can be used to generate a classification result by Eq. (9) calculated from Eq. (8). However, this paper proposes the SVC for data fusion, and hence the final step as given by Eq. (9) is removed from each GDBM. Instead, the outputs of the softmax (i.e., Eq. (8)) from all the constituting GDBMs are employed as the input vector $\mathbf{x}_n$ of the SVC (illustrated by Eq. (12)). Again, the training labels (i.e., $\hat{i}$) are applied for the SVC training. With the aforementioned updates, the SVC can be trained for the classification fusion of the MDSVC model.

With the trained MDSVC model, upon feeding a vibration measurement from the gearbox, the MDSVC model is capable of diagnosing the fault condition, and if any, the detailed fault pattern.

Having introduced construction, training and testing of the MDSVC approach separately, the procedure of applying the present MDSVC model for the gearbox fault diagnosis can be illustrated in Fig. 3 and is summarized as follows.

Step 1. Collect vibration signals $s(t)$, define fault patterns and diagnosis targets for a gearbox;
Step 2. Calculate time, frequency and wavelet modalities of the gearbox vibration measurement using Eq. (25);
Step 3. Construct three GDBMs according to Fig. 1 to accommodate three modalities, respectively;
Step 4. On the top of all the three GDBMs, construct an SVC to fuse the three GDBMs for developing the MDSVC model;

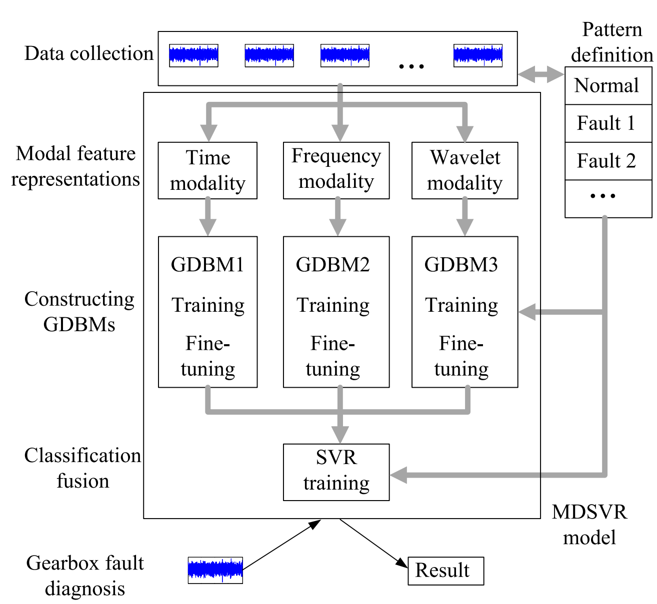


Fig. 3. Illustration of the present MDSVC model for the gearbox fault diagnosis.

Step 5. Pretrain and fine-tune the three constituting GDBMs separately using training dataset;
Step 6. Train the SVR using the training data given by Eq. (12) so as to complete the MDSVC training; and
Step 7. Diagnose the gearbox fault pattern from a vibration measurement using the trained MDSVC model.

4. Experiment evaluations

Two different configurations of a gearbox were used to carried out the experiments. Collected vibration signals were used for MDSVR model training and fault diagnosis testing, respectively. Experimental set-ups are introduced in the first subsection, followed by the results and discussions given in the second subsection.

4.1. Experimental configurations

The experiments were carried out on a gearbox fault diagnosis test-rig (fabricated by the lab of the Universidad Politécnica Salesiana, Ecuador) as shown in Fig. 4(a). A motor (SIEMENS, 3∼, 2.0HP) was connected to the input shaft of a gearbox via a coupling. The output shaft of the gearbox was connected with an electromagnetic torque break (ROSATI, 8.83 kW) through a belt transmission. The load of the torque break can be manually adjusted by a controller (TDK-Lambda, GEN 100-15-IS510). An accelerometer (PCB ICP 353C03) was mounted on the top of the gearbox to collect vibration signals, which were sent to a laptop (HP Pavilion g4-2055la) via a data acquisition box (DAQ, NI cDAQ-9234).

The gearbox was designed with two configurations: spur gearbox configuration (#1) and helical gearbox configuration (#2). As shown in Fig. 4(b), upon setting the gearbox at #1 configuration, two spur gears (number of teeth $Z_1=53$, and $Z_2=80$ with modulus 2.25 and impact angle 20°) were installed on the input and the output shafts of the gearbox, respectively. As for #2 configuration as shown in Fig. 4(c), the input gear and the output gear were chosen as helical gears (number of teeth $Z_1=30$, and $Z_4=80$ with modulus=2.25, impact angle 20°, and helical angle 20°). An intermediate shaft with two helical gears ($Z_2=Z_3=45$) was installed between the input and the output shafts for the transmission. Faulty components used for the spur gearbox and the helical gearbox are displayed in Fig. 4(d) and (e).

4.2. Fault diagnosis for spur gearbox

The proposed MDSVC approach was first tested with #1 configuration. During #1 experiment, in addition to normal gears, 5 faulty input gears and 3 faulty output gears (shown in Fig. 4(d)) were used to compose 10 condition patterns as listed in Table 1. In this experiment, 3 different load conditions were set for each pattern. For each pattern and load condition, the tests were repeated 5 times. In each time of the test, 24 signals each of which with duration 0.4096 s were collected, with a sampling frequency of 10 kHz. In this way, we obtained 3600 vibration signals corresponding to 10 condition patterns for #1 experiment.

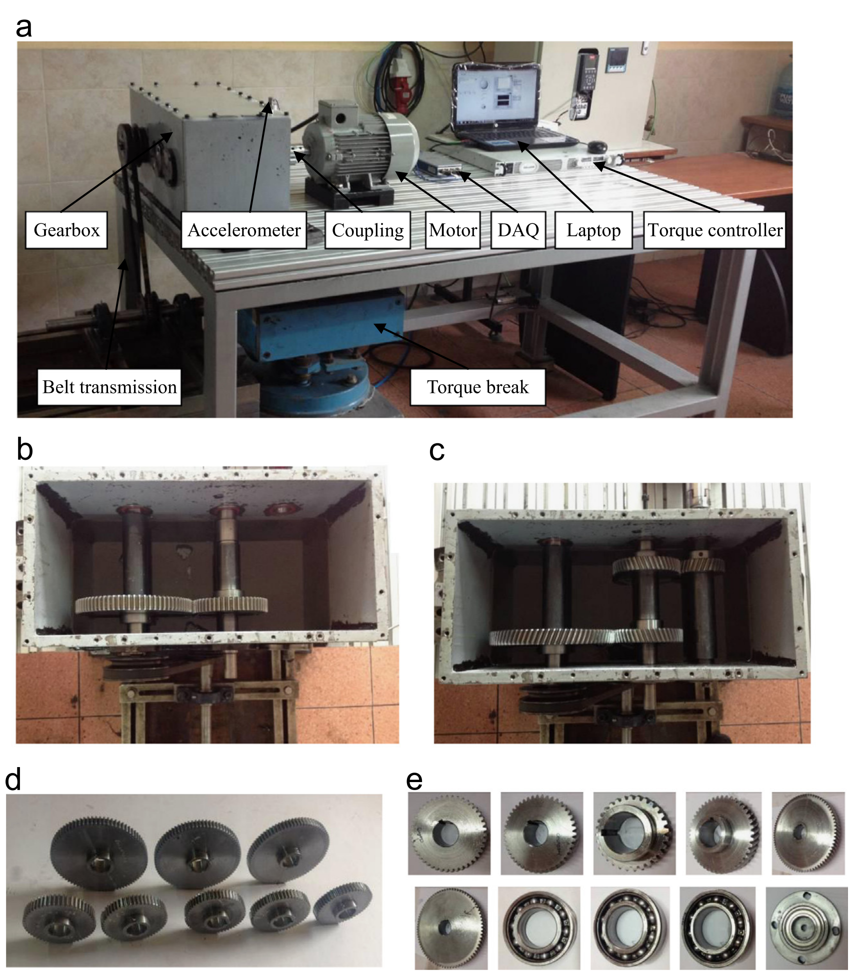


Fig. 4. Experimental configurations: (a) experimental set-up; (b) spur gearbox configuration; (c) helical gearbox configuration; (d) faulty components for the spur gearbox experiment; and (e) faulty components for the helical gearbox experiment.

Through disordering all the recorded signals, 2400 signals were used for the MDSVC model training, while the rest 1200 ones for the fault diagnosis testing. The MDSVC model was developed with parameters as: three GDBMs for time, frequency and wavelet modalities, respectively; for each GDBM, number of hidden layers = 2, number of neurons in the hidden layer 1 = 50, number of neurons in the hidden layer 2 = 50, number of pretraining epochs = 50, number of fine-tuning epochs = 250, initial learning rate = 0.001, its upper-bound = 0.001, and weight decay = 0.005; and input dimension of the SVR = 30.

The trained MDSVC model was then used for spur gearbox fault diagnosis. For 1200 testing signals, the present approach classified 1165 signals with correct patterns. In other words, the classification rate of the present MDSVC model is 97.08% for the spur gearbox fault diagnosis. For comparison, the GDBM and the SVC models were applied for the fault diagnosis with time, frequency, and wavelet modalities, and the combination of the three modalities, respectively. For a fair comparison, all the peer models were trained using the same parameters (if applicable) with the MDSVC. Comparison of results is displayed in Table 2.

From spur gearbox diagnosis results as shown in Table 2, clearly the classification rate for the present MDSVC model is the best (97.08%) comparing to the GDBM and the SVC ones. It is noting that in this work we have not employed more deep or shallow learning models such as the deep believe networks, the deep autoencoders, the decision tree, and the neural networks. On the one hand, the SVR has been proven the prominent model which outperforms most of shallow learning members. On the other hand, the GDBM is one of the popular deep learning representatives. Hence the GDBM and the SVC should be at the top level in the classification community. Comparing to these two state-of-the-art methods, the MDSVC features evident improvement in terms of fault classification rate. This indicates that the deep support vector classification with multimodal homologous features is capable of improving the gearbox fault diagnosis performance.

4.3. Fault diagnosis for helical gearbox

To validate the robustness of the present MDSVC approach, we also carried out the experiment with #2 configuration. In this experiment, 6 faulty gears (1 input gear, 3 intermediate ones and 2 output ones), 3 faulty bearings (with 1 inner race fault, 1 outer race fault and 1 ball fault, respectively), and 1 eccentric bearing box were used to form 11 condition patterns as listed in Table 3.

Similar with the spur gearbox diagnosis experiment, 3 different load conditions, 5 repetitions, 24 signals each of which with duration 0.4096 s, and 10 kHz sampling frequency were set in #2 experiment. Total number of vibration signals is 3960. Among them 2/3 numbers were random chosen for modeling, and the rests for testing. The MDSVC model was again constructed with the same parameters used in the spur gearbox diagnosis experiment. With the trained MDSVC model, 1320 testing signals were used for performance validation, shown in Table 4. Reference models, i.e., the GDBM and the SVC models, were again employed for the fault diagnosis with different input modalities. Comparison of results is also included in Table 4.

According to the helical gearbox diagnosis results as shown in Table 4, it is shown that the present MDSVC model again outperforms peer models, no matter what input modality is applied. This means that the present MDSVC approach is robust for the gearbox fault classifications with different types.

It should be noted that according to the research [43], better classification rates can be obtained by the deep learning with "deeper" network parameters. In this work, all deep learning models are under small scale training for comparison. With more pretraining and fine-tuning epochs, the trained MDSVC model can achieve better classification rate for the gearbox fault diagnosis, but needing more training computational resources.

Table 1
Condition patterns of the #1 experiment.

Pattern number	Input gear Z_1	Output gear Z_2	Load
A	Normal	Normal	Zero, small, great
B	Chaffing tooth	Normal	Zero, small, great
C	Worn tooth	Normal	Zero, small, great
D	Chipped tooth 25%	Normal	Zero, small, great
E	Chipped tooth 50%	Normal	Zero, small, great
F	Missing tooth	Normal	Zero, small, great
G	Normal	Chipped tooth 25%	Zero, small, great
H	Normal	Chipped tooth 50%	Zero, small, great
I	Normal	Missing tooth	Zero, small, great
J	Chipped tooth 25%	Chipped tooth 25%	Zero, small, great

Table 2
Spur gearbox fault classification rates for the testing dataset.

Model	Input modality			
	Time	Frequency (%)	Wavelet (%)	Three modalities (%)
MDSVC	/	/	/	97.08
GDBM	35.00	78.08	90.92	92.08
SVC	60.83	52.83	69.50	54.75

Table 3
Condition patterns of the #2 experiment.

Pattern number	Faulty component	Fault detail	Load
1	N/A	N/A	Zero, small, great
2	Input gear Z_1	Worn tooth	Zero, small, great
3	Intermediate gear Z_2	Chaffing tooth	Zero, small, great
4	Intermediate gear Z_3	Pitting tooth	Zero, small, great
5	Intermediate gear Z_3	Worn tooth	Zero, small, great
6	Output gear Z_4	Chipped tooth	Zero, small, great
7	Output gear Z_4	Chaffing tooth	Zero, small, great
8	Bearing of output shaft	Inner race fault	Zero, small, great
9	Bearing of output shaft	Outer race fault	Zero, small, great
10	Bearing of output shaft	Ball fault	Zero, small, great
11	Bearing box of output shaft	Eccentric	Zero, small, great

Table 4
Helical gearbox fault classification rates for the testing dataset.

Model	Input modality			
	Time (%)	Frequency (%)	Wavelet (%)	Three modalities (%)
MDSVC	/	/	/	88.41
GDBM	36.21	54.24%	82.88	73.18
SVC	48.11	35.00	79.24	82.20

5. Conclusions

In this paper, a multimodal deep support vector classification (MDSVC) model has been developed to diagnose gearbox faults. This model employed the separation–fusion based deep learning strategy to tackle the fault pattern classification problem. On the one hand, to understand gearbox healthy condition patterns in a "wide" fashion, time, frequency and wavelet modalities of vibration signals were generated as homologous features. On the other hand, to represent the gearbox condition in a "deep" style, the Gaussian-Bernoulli deep Boltzmann machine (GDBM) was applied to learn features in each modality. Moreover, the support vector classification (SVC) was used to fuse three modalities and their deep learning representations in an integrated MDSVC model. Comparing with the GDBM and the SVC models, the proposed MDSVC is capable of obtaining the best diagnosis performance, due to simultaneous exploration of the wide and deep aspects of gearbox conditions.

It should be pointed out that the homologous features were used for all the modalities in this paper. This means that with only one vibration source can the MDSVC obtain good fault diagnosis result for the gearbox. If different measurements such as vibration, acoustic and thermal signals are available, the proposed MDSVC model can be used as a multi-source data fusion approach. In the future work, much attention should be paid on the application of the MDSVC method to multiple sensor fusion in fault diagnosis applications.

Acknowledgements

This work is supported in part by the Prometeo Project of the Secretariat for Higher Education, Science, Technology and Innovation of the Republic of Ecuador, the National Natural Science Foundation of China (51375517), the Project of Chongqing Innovation Team in University (KJTD201313), and the Youth Science and Technology Talents Project of Chongqing (cstc2014kj-rc-qnrc00003). The valuable comments and suggestions from the editors and the two reviewers are very much appreciated.

References

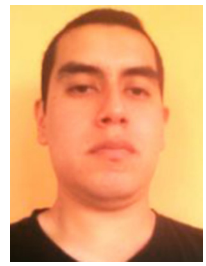
[1] H.H. Bafroui, A. Ohadi, Application of wavelet energy and Shannon entropy for feature extraction in gearbox fault detection under varying speed conditions, Neurocomputing 133 (2014) 437–445.
[2] Y. Lei, J. Lin, M.J. Zuo, Z. He, Condition monitoring and fault diagnosis of planetary gearboxes: a review, Measurement 48 (2014) 292–305.
[3] Z. Su, B. Tang, Z. liu, Y. Qin., Multi-fault diagnosis for rotating machinery based on orthogonal supervised linear local tangent space alignment and least square support vector machine, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2015.01.016.
[4] C. Li, M. Liang, Separation of vibration-induced signal of oil debris sensor for vibration monitoring, Smart Mater. Struct. 20 (3) (2011) 045019.
[5] D.H. Pandya, S.H. Upadhyay, S.P. Harsha, Fault diagnosis of rolling element bearing with intrinsic mode function of acoustic emission data using APF-KNN, Expert Syst. Appl. 40 (10) (2013) 4137–4145.
[6] C. Li, M. Liang, Time–frequency signal analysis for gearbox fault diagnosis using a generalized synchrosqueezing transform, Mech. Syst. Signal Process. 26 (1) (2012) 205–217.
[7] S. Yin, Z. Huang, Performance monitoring for vehicle suspension system via fuzzy positivistic *C*-means clustering based on accelerometer measurements, IEEE/ASME Trans. Mechatron. (2015), http://dx.doi.org/10.1109/TMECH.2014.2358674.
[8] J.D. Wu, C.Q. Chuang, Fault diagnosis of internal combustion engines using visual dot patterns of acoustic and vibration signals, NDT E Int. 38 (8) (2005) 605–614.
[9] C. Li, M. Liang, T. Wang, Criterion fusion for spectral segmentation and its application to optimal demodulation of bearing vibration signals, Mech. Syst. Signal Process 64–65 (2015) 132–148.
[10] M. Amarnath, I.R.P. Krishna, Local fault detection in helical gears via vibration and acoustic signals using EMD based statistical parameter analysis, Measurement 58 (2014) 154–164.
[11] M.H. Gharavian, F.A. Ganj, A.R. Ohadi, H.H. Bafroui, Comparison of FDA-based and PCA-based features in fault diagnosis of automobile gearboxes, Neurocomputing 121 (2013) 150–159.
[12] R. Yan, R.X. Gao, X. Chen, Wavelets for fault diagnosis of rotary machines: a review with applications, Signal Process. 96 (2014) 1–15.
[13] K. Noda, H. Arie, Y. Suga, T. Ogata., Multimodal integration learning of robot behavior using deep neural networks, Rob. Autom. Syst. 62 (6) (2014) 721–736.
[14] Y. Wang, G. Xu, L. Liang, K. Jiang, Detection of weak transient signals based on wavelet packet transform and manifold learning for rolling element bearing fault diagnosis, Mech. Syst. Signal Process. 54–55 (2015) 259–276.
[15] A. Raad, J. Antoni, M. Sidahmed, Indicators of cyclostationarity: theory and application to gear fault monitoring, Mech. Syst. Signal Process. 22 (3) (2008) 574–587.
[16] C. Li, M. Liang, Continuous-scale mathematical morphology-based optimal scale band demodulation of impulsive feature for bearing defect diagnosis, J. Sound Vib. 331 (2012) 5864–5879.
[17] C. Shen, D. Wang, F. Kong, P.W. Tse, Fault diagnosis of rotating machinery based on the statistical parameters of wavelet packet paving and a generic support vector regressive classifier, Measurement 46 (2013) 1551–1564.
[18] Y. Lei, D. Kong, J. Lin, M.J. Zuo, Fault detection of planetary gearboxes using new diagnostic parameters, Meas. Sci. Technol. 23 (2012) 055605.
[19] S.S. Tayarani-Bathaie, Z.N.S. Vanini, K. Khorasani, Dynamic neural network-based fault diagnosis of gas turbine engines, Neurocomputing 125 (11) (2014) 153–165.
[20] L. Guo, J. Chen, X. Li., Rolling bearing fault classification based on envelope spectrum and support vector machine, J. Vib. Control 15 (9) (2009) 1349–1363.
[21] F. Chen, B. Tang, R. Chen., A novel fault diagnosis model for gearbox based on wavelet support vector machine with immune genetic algorithm, Measurement 46 (1) (2013) 220–232.
[22] S. Yin, S.X. Ding, X. Xie, H. Luo., A review on basic data-driven approaches for industrial process monitoring, IEEE Trans. Ind. Electron. 61 (11) (2014) 6418–6428.
[23] S. Yin, X. Li, H. Gao, O. Kaynak, Data-based techniques focused on modern industry: an overview, IEEE Trans. Ind. Electron. 62 (1) (2015) 657–667.
[24] G.E. Hinton, R.R. Salakhutdinov, Reducing the dimensionality of data with neural networks, Science 313 (2006) 504–507.
[25] S. Zhou, Q. Chen, X. Wang, Fuzzy deep belief networks for semi-supervised sentiment classification, Neurocomputing 131 (2014) 312–322.
[26] B. Leng, X. Zhang, M. Yao, Z. Xiong., A 3D model recognition mechanism based on deep Boltzmann machines, Neurocomputing 151 (2015) 593–602.
[27] K. Katayama, M. Ando, T. Horiguchi, Model of MT and MST areas using an autoencoder, Physica A: Stat. Mech. Appl. 322 (2013) 531–545.
[28] T.N. Sainath, B. Kingsbury, G. Saon, H. Soltau, A. Mohamed, G. Dahl, B. Ramabhadran, Deep convolutional neural networks for multi-modality isointense infant brain image segmentation, NeuroImage 108 (2015) 214–224.
[29] V.T. Tran, F.A. Thobiani, A. Ball., An approach to fault diagnosis of reciprocating compressor valves using Teager–Kaiser energy operator and deep belief networks, Expert Syst. Appl. 41 (2014) 4113–4122.
[30] P. Tamilselvan, P. Wang, Failure diagnosis using deep belief learning based health state classification, Reliab. Eng. Syst. Saf. 115 (2013) 124–135.
[31] S. Yin, X. Zhu, O. Kaynak, Improved PLS focused on key-performance-indicator-related fault diagnosis, IEEE Trans. Ind. Electron. 62 (3) (2015) 1651–1658.
[32] S. Yin, G. Wang, X. Yang, Robust PLS approach for KPI-related prediction and diagnosis against outliers and missing data, Int. J. Syst. Sci. 45 (7) (2014) 1375–1382.
[33] R. Salakhutdinov, G. Hinton., An efficient learning procedure for deep Boltzmann machines, Neural Comput. 24 (8) (2012) 1967–2006.
[34] K.H. Cho, T. Raiko, A. Ilin, Gaussian-Bernoulli deep Boltzmann machine, in: The International Joint Conference on Neural Networks (IJCNN), 2013, pp. 1–7.
[35] N. Srivastava, R. Salakhutdinov., Multimodal learning with deep Boltzmann machines, J. Mach. Learn. Res. 15 (2014) 2949–2980.
[36] S. Yin, X. Zhu, C. Jing, Fault detection based on a robust one class support vector machine, Neurocomputing 145 (2014) 263–268.
[37] Y. Tian, M. Fu, F. Wu, Steel plates fault diagnosis on the basis of support vector machines, Neurocomputing 151 (2015) 296–303.
[38] Y. Tang, Deep learning using linear support vector machines, in: Workshop on Representational Learning, ICML 2013, Atlanta, USA, 2013.
[39] I.S. Bozchalooi, M. Liang, A smoothness index-guided approach to wavelet parameter selection in signal de-noising and fault detection, J. Sound Vib. 308 (2007) 246–267.
[40] J. Rafiee, M.A. Rafiee, P.W. Tse, Application of mother wavelet functions for automatic gear and bearing fault diagnosis, Expert Syst. Appl. 37 (2010) 4568–4579.
[41] D. Hjelma, V. Calhouna, R. Salakhutdinov, E. Allena, T. Adali, S. Plisa, Restricted Boltzmann machines for neuroimaging: an application in identifying intrinsic networks, NeuroImage 96 (2014) 245–260.
[42] R. Salakhutdinov, Learning Deep Generative Models Ph.D., Thesis, University of Toronto, 2009.
[43] J. Schmidhuber, Deep learning in neural networks: an overview, Neural Netw. 61 (2015) 85–117.

Chuan Li (chuanli@21cn.com) received his Ph.D. degree from the Chongqing University, China, in 2007. He has been successively a Postdoctoral Fellow with the University of Ottawa, Canada, a Research Professor with the Korea University, South Korea, and a Senior Research Associate with the City University of Hong Kong, China. He is currently a Professor with the Chongqing Technology and Business University, China, and a Prometeo Researcher with the Universidad Politécnica Salesiana, Ecuador. His research interests include machinery healthy maintenance, and intelligent systems.

René-Vinicio Sanchez (rsanchezl@ups.edu.ec) received the B.S. in Mechanical Engineering in 2004, from the Universidad Politécnica Salesiana (UPS), Ecuador. He got his master in management audit quality in 2008 at the UTPL, Ecuador, and the master degree in industrial technologies research in 2012 at the UNED, Spain. Currently, he is Professor of the Department of Mechanical Engineering in the UPS. His Research interests are in machinery health maintenance, pneumatic and hydraulic systems, artificial intelligence and engineering education.

Grover Zurita (gzuritav@ups.edu.ec) received his Ph.D. degree from Luleå University of Technology, Sweden, in 2001. He was a Postdoctoral Fellow at the University of New South Wales, Australia, in 2002. Currently, he is a Professor at the Private University of Bolivia, and a Prometeo Researcher at the Universidad Politécnica Salesiana of Ecuador. His research interests are machine diagnosis, optimization and control of internal combustion engines.

Mariela Cerrada (cerradam@ula.ve) received her Ph.D. degree in Automatic Systems in 2003 from the INSA Toulouse, France. She is currently a full dedicated time titular professor in the Department of Control Systems and associate member of the Studies Center on Microcomputers and Distributed Systems (CEMISID) at the Engineering Faculty in the Universidad de Los Andes of Venezuela, and a Prometeo Researcher in the Universidad Politécnica Salesiana of Ecuador. Her main research area is on fault diagnosis, supervision and intelligent control systems.



Diego Cabrera (dcabrera@ups.edu.ec) received his M. Sc. degree at the Sevilla University in 2014. Currently, he is a lecturer at the Universidad Politécnica Salesiana (UPS), Ecuador. He is a member of the research group of innovation and development at UPS, and a member of the research group of complex systems modelling at the Universidad Central, Ecuador. His research areas are condition-based maintenance, complex systems modelling, and intelligence systems.

Rafael E. Vásquez (rafael.vasquez@upb.edu.co) received his Ph.D. in Mechanical Engineering from the University of Florida, USA, in 2011. He is currently Full Professor in the Department of Mechanical Engineering at the Universidad Pontificia Bolivariana, in the area of dynamics, systems and control. His research interests are theory of mechanisms; robotics; design, analysis, and control of dynamic systems; tensegrity systems; new technologies for energy harvesting; machine monitoring; and engineering education.