Journal of Physics: Conference Series

PAPER • **OPEN ACCESS**

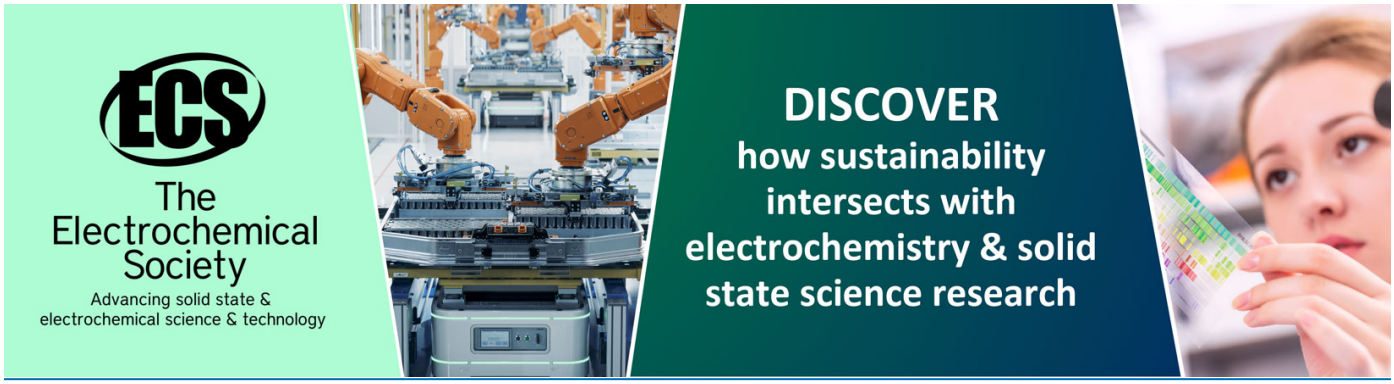
Topic Correlation Analysis for Bearing Fault Diagnosis Under Variable Operating Conditions

To cite this article: Chao Chen *et al* 2017 *J. Phys.: Conf. Ser.* **842** 012045

View the article online for updates and enhancements.

You may also like

- Domain adaptation with domain specific information and feature disentanglement for bearing fault diagnosis
Shaozhang Xie, Peng Xia and Hanqi Zhang
- A working-condition-robust method for two-stage health measurement of rolling bearings based on energy entropy distribution and dynamic adversarial transfer network
Wei Jiang, Xiaoming Xue, Nan Zhang et al.
- Fault diagnosis of rolling bearings based on dynamic convolution and dual-channel feature fusion under variable working conditions
Dechen Yao, Tao Zhou, Jianwei Yang et al.



This content was downloaded from IP address 3.137.157.45 on 04/05/2024 at 05:15

Topic Correlation Analysis for Bearing Fault Diagnosis Under Variable Operating Conditions

Chao Chen[1], Fei Shen[1] and Ruqiang Yan[1,*]

[1]School of Instrument Science and Engineering, Southeast University
Nanjing, Jiangsu 210096, China

[*]Email: ruqiang@seu.edu.cn

Abstract. This paper presents a Topic Correlation Analysis (TCA) based approach for bearing fault diagnosis. In TCA, Joint Mixture Model (JMM), a model which adapts Probability Latent Semantic Analysis (PLSA), is constructed first. Then, JMM models the shared and domain-specific topics using "fault vocabulary". After that, the correlations between two kinds of topics are computed and used to build a mapping matrix. Furthermore, a new shared space spanned by the shared and mapped domain-specific topics is set up where the distribution gap between different domains is reduced. Finally, a classifier is trained with mapped features which follow a different distribution and then the trained classifier is tested on target bearing data. Experimental results justify the superiority of the proposed approach over the stat-of-the-art baselines and it can diagnose bearing fault efficiently and effectively under variable operating conditions.

1. Introduction

Rolling bearings, as one of the most important components in rotary machines, are widely used in modern manufacturing systems. However, faults occurring in bearings could cause dreadful consequences such as performance deterioration, costly downtime, production delay or even personnel casualties [1]. Therefore, detecting the faults occurring in bearings as early as possible can not only ensure reliability, safety and good operation of mechanical equipment, but also have the ability to avoid unnecessary losses. With the development of fault diagnosis in rolling bearings, various methods [2-4] have been employed to measure certain physical quantities, including temperature monitoring, oil analysis, vibration signal analysis, and etc., which serve as a crucial preprocessing step for fault diagnosis. Among them, the use of vibration signal analysis has been proven to be quite common, practical and effective. In many existing researches [5-6], the fault diagnosis of bearings can be treated as a problem of pattern recognition solved by machine learning-based algorithms using extracted features from vibration data. However, in many applications, bearings are operated under complex working conditions [7], such as high speed, variable load, high temperature and other complex conditions, effective fault diagnosis for bearings is challenging.

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

Nowadays, most studies of bearing fault diagnosis focus on a specific working condition(e.g., in laboratory or under controlled conditions), where traditional machine learning methods (e.g., back propagation neural network, transductive support vector machine, K-means) are proved to be effective using labeled, unlabeled data or both of them[8-10]. However, successful utilization of all these methods requires that the training and testing data are drawn from the same distribution (e.g., vibration data for training and testing process are measured from the same set of experiments) [11]. In many real-world applications, bearings are working under variable operating conditions, where there exist various speeds, changing load as well as some unknown jamming, there is no guarantee that the training and testing data are drawn from the same distribution. When the distribution changes, it is common to rebuild the diagnostic model using sufficient newly labeled data. But it is sometimes time-consuming or expensive to obtain vibration data of the target bearing to train a precise classifier, because the installation position of bearing is inside some unable dismantled parts, thus the performance of traditional machine learning methods in terms of bearing fault diagnosis under complex working conditions would come out to be poor. In contrast, transfer learning, a new machine learning approach with ability to recognize and apply knowledge and skills learned in previous tasks to new tasks, has attracted more and more attention since 1995 [12-13]. It is a good solution for fault diagnosis in complex situations, where the training and test data are drawn from different distributions.

Inspired by this, the paper proposes a novel bearing fault diagnosis strategy based on Topic Correlation Analysis (TCA). In TCA, Joint Mixture Model (JMM), a model which adapts Probability Latent Semantic Analysis (PLSA), is constructed first. As JMM is originally designed well for text classification, the following modifications have been taken to ensure success in applying JMM to fault diagnosis: (1) view a time series of vibration data as a document consisted of a number of “fault word”, and take energy distribution as well as load, speed, and other parameters related to vibration data as the topics embedded in the document; (2) extract frequency-domain features and time-domain features from vibration data; (3) apply K-means to generate a set of “fault word” using extracted features. Then, JMM models the shared and domain-specific topics using “fault word”. After that, the correlations between two kinds of topics are computed and used to construct a mapping matrix. Furthermore，a new shared space spanned by the shared and mapped domain-specific topics is built where the distribution gap between different domains is reduced. Finally, a classifier is trained with mapped features which follow a different distribution and then the trained classifier is tested on target bearing.

The rest of the paper is organized as follows. In section 2, the principle of Topic Correlation Analysis (TCA) along with its modification, applied in bearing fault diagnosis, is introduced. Then experimental study using bearing data set is carried out in Section 3, followed by the conclusions in Section 4.

2. TCA-based transfer learning approach for bearing fault diagnosis

In practice, classification algorithm based on machine learning plays an important role in fault diagnosis. But a basic assumption of traditional machine learning is that the training and testing data follow the same distribution. However, when it comes to bearing fault diagnosis under variable conditions, the distributions of training and testing data do not always satisfy the assumption. In this case, transfer learning will become an ideal tool for bearing fault diagnosis.

Here an improved algorithm, named enhanced TCA, which can reduce the data difference between the two data samples, is put forward for fault diagnosis. TCA is based on the model of Probabilistic Latent Semantic Analysis (PLSA) [14], and the principle of PLSA is introduced first.

2.1. A Brief Review of PLSA

PLSA, originating from latent semantic analysis (LSA), has been widely used for document clustering and related tasks. Given a set of documents $D=\{d_1,d_2,\ldots,d_N\}$, words $W=\{w_1,w_2,\ldots,w_M\}$ and latent topic variables $Z=\{z_1,z_2,\ldots,z_K\}$. A generative model for document co-occurrences, as is shown in Figure 1, can be defined by the following scheme:

- select a document d_i with probability $P(D = d_i)$,
- draw a topic z_k with probability $P(Z = z_k | D = d_i)$,
- select a word w_j with probability $P(W = w_j | Z = z_k)$.

Where $P(d_i)$ denotes the probability that a word occurrence will be observed in a particular document d_i , $P(w_j | z_k)$ denotes the class-conditional probability of a specific word conditioned on latent topic variable z_k , and $P(w_j | d_i)$ represents a document specific probability distribution over the latent variable space.

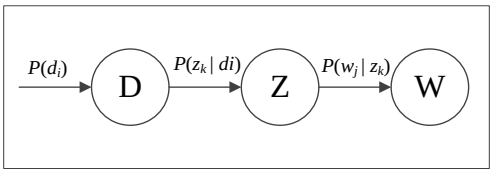


Figure 1. Graphical representation of Probabilistic Latent Semantic Analysis (PLSA) model.

Then $P(d_i, w_j)$ can be obtained according to Bayesian formula

$$P(d_i,w_j)= P(d_i)\, P(w_j | d_i) \tag{1}$$

Where

$$P(w_j | d_i) = \sum_{k=1}^{K} P(z_k | d_i) P(w_j | z_k) \tag{2}$$

Then, the Expectation-Maximization (EM) algorithm, which is suitable for estimating model parameters of latent variable model [15], is applied. The aim of EM is to find suitable $P(z_k | d_i)$ and $P(w_j | z_k)$ by maximizing objective function

$$L = \sum_{i=1}^{N} \sum_{j=1}^{M} n\left(d_i, w_j\right) \log P\left(d_i, w_j\right) \tag{3}$$

where $n(d_i, w_j)$ denotes the number of times that word w_j occurs in document d_i.

PLSA model is originally designed to address the semantic problem caused by synonyms and polysemy in document clustering task. Similarly, in the field of fault diagnosis, fault features, equivalent to words in document categorization, also have the characteristics of synonyms and polysemy. For example, there must be a number of commonalities between features extracted from the same fault type under different working conditions because of the same root cause of failures, which is similar to synonyms in text. At the same time, due to these commonalities, the extracted features can not only represent their own fault category, but also have a certain contribution to the clustering of the fault type under another working condition, under this circumstance, these fault features can also be seen as polysemy.

In document clustering task, vectors which are represented by word frequency are used to stand for documents. The latent 'bag-of-words' is used to discovery the topics of documents. Thus, in fault diagnosis, if one fault type is viewed as a document consisted of a number of "fault word", and energy distribution as well as load, speed, and other parameters related to vibration data is taken as the latent topics embedded in the document, then the probability distribution of the latent topic can be found with PLSA. Specifically, given a set of vibration signal $D=\{d_1,d_2,\ldots,d_N\}$, some modifications are listed as follows:

(1) Extracting time-domain features and frequency domain features respectively from the vibration signal, then utilizing k-means to generate N fault words

(2) Computing the similarity between each fault word and the features extracted from each vibration signal, then co-occurrence matrix $n(d_i, w_j)$ is obtained, where $n(d_i, w_j)$ denotes the similarity between fault word w_j and vibration signal d_i.

Then, the rest steps are the same as traditional PLSA used in document clustering task.

2.2. TCA model

Before discussing TCA, some basic notations are listed in Table 1.

Table 1. Notations

Symbols	Description
l	domain random variable, $l \in \{t, s\}$
D^l	fault type random variable
Z	topic random variable
π	decision random variable, $\pi \in \{0,1\}$
W	feature word of fault information
d_n^l	nth fault type in domain l
w_j	jth fault word
z_k	kth shared topic
z_r^l	rth domain-specific topic in domain l
$\mu_{d_i^l}$	fault type-level parameter for the decision process
K	number of the shared topics
K^l	number of the domain-specific topics in domain l

The main idea of TCA can be summarized into three parts:

Part I. Modelling both Shared and Domain-Specific Topics

In this part, Joint Mixture Model(JMM), which adapts PLSA, is used to extracting both shared and domain specific topics. In JMM, as is shown in Figure 2 (a), the co-occurrence between a fault sample and a fault word is associated with a latent topic $Z \in \{z_1,...,z_K, z_1^s,...,z_{K^s}^s, z_1^t,...,z_{K^t}^t\}$ which is drawn from either the shared topics, i.e. $\{z_1,...,z_K\}$, or the domain-specific topics, i.e. $\{z_1^s,...,z_{K^s}^s, z_1^t,...,z_{K^t}^t\}$. A latent decision variable $\pi \in \{0,1\}$ is used to control which kind of topic is drawn. The generative process for co-occurrence in JMM is defined as:

(1) Selecting a kind of fault sample d_i with probability $P(D = d_i)$

(2) Drawing a decision random variable $\pi \sim \text{Bern}(\mu_{d_i^l})$:if $\pi = 0$, a shared topic z_K with probability $P(Z = z_k | D = d_i, \pi=0)$ is chose, and a fault word w_j with probability $P(W = w_j | Z = z_k)$ is generated; if $\pi = 1$, a domain-specific topic Z_r^l with probability $P(Z = z_r^l | D^l = d_i^l, \pi = 1)$ is chosen, and a fault word w_j with probability $P(W = w_j | Z = z_r^l)$ is generated.

(3) The joint probability model for the data generative process is described as:

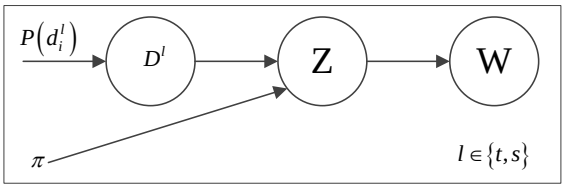
$$P(d_i^l, w_j) = P(d_i^l)P(w_j | d_i^l) \tag{4}$$

where

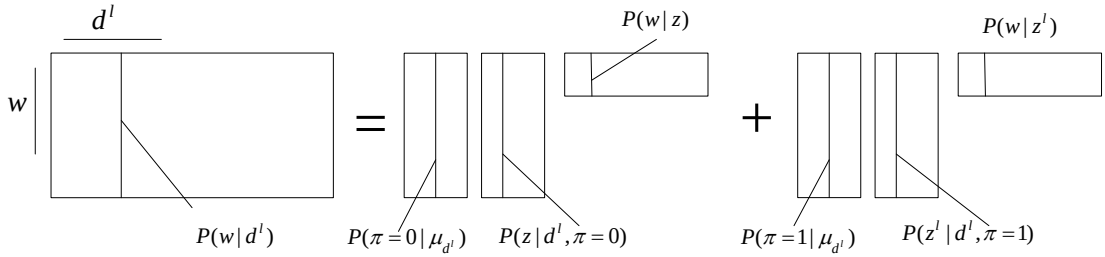
$$\begin{aligned} P(w_j | d_i^l) &= P(\pi = 0 | \mu_{d_i^l}) \sum_{k=1}^{K} P(z_k | d_i^l, \pi = 0) P(w_j | z_k) \\ &+ P(\pi = 1 | \mu_{d_i^l}) \sum_{r=1}^{K^l} P(z_r^l | d_i^l, \pi = 1) P(w_j | z_r^l) \end{aligned} \tag{5}$$

The equation (5) related matrix decomposition is illustrated in Figure 2 (b). Then EM algorithm is used to estimate model parameters of the latent variable model by maximizing following objective function

$$L = \sum_{l \in \{s,t\}} \sum_{d_i^l \in D^l} \sum_{w_j \in W} n(d_i^l, w_j) \log P(d_i^l, w_j) \tag{6}$$



(a)



(b)

Figure 2 : (a) JMM graphical model; **(b)** Diagram of matrix decomposition

Part II. Measuring Topic Correlations across Domains

Correlations between domain-specific topics from different domains can be measured as follows:

(1) Calculating the similarity $\theta_{z_i^l, z_k}$ between the shared and the domain specific topics in each domain with the vectorial angle cosine method

(2) The Pearson's Correlation Coefficients (PCCs) [16] is adopted to infer the correlations between the domain-specific topics from different domains. With PCCs, a topic mapping matrix $U = \left\{\rho\left(z_i^t, z_j^s\right)\right\}$, where $i = 1,...,K^t$, $j = 1,...,K^s$, $U \in R^{K^t \times K^s}$, can be obtained as

$$\rho\left(z_i^t, z_j^s\right) = \frac{\sum_k \left(\theta_{z_i^t, z_k} - \bar{\theta}_{z_i^t}\right)\left(\theta_{z_j^s, z_k} - \bar{\theta}_{z_j^s}\right)}{\sqrt{\sum_l \left(\theta_{z_i^t, z_k} - \bar{\theta}_{z_i^t}\right)^2}\sqrt{\sum_l \left(\theta_{z_j^s, z_k} - \bar{\theta}_{z_j^s}\right)^2}} \tag{7}$$

where $\bar{\theta}_{z_i^l}$ represents the mean similarity between z_i^l and all the shared topics.

Part III. Constructing New Representation for Documents

Here both shared and the mapped domain-specific topics are utilized to construct a new shared space, which can be described as follows:

(1) Representing fault samples $\left\{d_i^l\right\}_{l,i}$ from both domains with the extracted topics:

$$\phi(d_i^l) = \left[P\left(z_1 | d_i^l\right),...,P\left(z_K | d_i^l\right),P\left(z_1^l | d_i^l\right),...,P\left(z_{K^l}^l | d_i^l\right)\right]^T \tag{8}$$

where $\phi(d_i^l)_{[1:K]}$ represents domain-shared part, and $\phi(d_i^l)_{\left[K+1:K^l\right]}$ stands for domain-specific part.

(2) Transforming the domain-specific part $\phi(d_i^t)_{\left[K+1:K+K^t\right]}$ with the topic mapping matrix U:

$$\varphi\left(\phi\left(d_i^t\right)\right) = \left[\phi\left(d_i^l\right)_{[1:K]}; U^T \phi\left(d_i^t\right)_{\left[K+1:K+K^t\right]}\right] \tag{9}$$

The new representation $\varphi\left(\phi\left(d_i^t\right)\right)$ is composed of shared topics and mapped source domain-specific topics. Then, each fault sample in the target domain is mapped into the new feature space spanned by $\varphi\left(\phi\left(d_i^t\right)\right)$.

2.3. Overview of Bearing Fault Diagnosis Based on Transfer Learning

The enhanced TCA algorithm can map the feature data that under different working conditions to a new shared space, where the difference between the data will be reduced. Therefore, the bearing fault diagnosis problem under variable conditions can be solved using this transfer learning technology. The complete process of the proposed method is summarized in Algorithm 1.

Algorithm 1: Bearing Fault Diagnosis Based on Enhanced TCA

Input: (1) Source domain labeled vibration signal and target domain unlabeled vibration signal; (2)Number of shared topics K; (3) Number of source/target domain-specific topics K^s, K^t;(4) Number of iterations T
Output: The predicted class label of each target unlabeled vibration signal
Extract time-domain features and frequency-domain features from inputted vibration signal.
Apply K-means to generate a set of "fault word" from the extracted features.
Initialize the parameters of JMM model;
for t←1 to T do
    EM algorithm
end
Measure the topic correlations;
Construct the topic mapping matrix U;
Represent each fault sample with extracted topics;
Transform each fault sample in the target domain;
Train a classifier with the labeled source domain data $D^s = \left\{\left(\phi\left(d_n^s\right), y_n^s\right)\right\}_{n=1}^{N^s}$ and predict the class labels of unlabeled target domain data $D^t = \left\{\left(\phi\left(d_i^t\right)\right)\right\}_{i=1}^{N^t}$

3. Experimental study and performance analysis

3.1. Experimental setup

To verify the effectiveness of the presented approach in bearing fault diagnosis, the bearing data set contributed by Western Reserve University Bearing Data Center was analyzed. As is shown in Figure 3, the experiment system consists of a 2 hp motor (left), a torque transducer/encoder (center), a dynamometer (right), and control electronics (not shown). The model of drive end bearing is SKF6205, and the model of fan end bearing is SKF6203.

Vibration signals were measured using a 16 channel DAT recorder with12 kHz sample frequency from the test system under variable speeds and loading conditions. Data were collected at 12,000 samples per second. Speed and horsepower data were collected using the torque transducer/encoder and were recorded by hand. Different bearing working conditions are listed in Table 2.

Table 2. Bearing test condition descriptions.

Experiment Condition	**1**	**2**	**3**	**4**	
Motor Speed (RPM)/ Motor Load (HP)	1797/ 0	1772/ 1	1750/ 2	1730/ 3	——
Fault Diameter (mm)	0.1778	0.3556	0.5334	0.7112	——
Fault Type	Inner Race	Ball	Outer Race Relative to Load Zone (Load Zone Centered at 6:00)		
			6:00	3:00	12:00

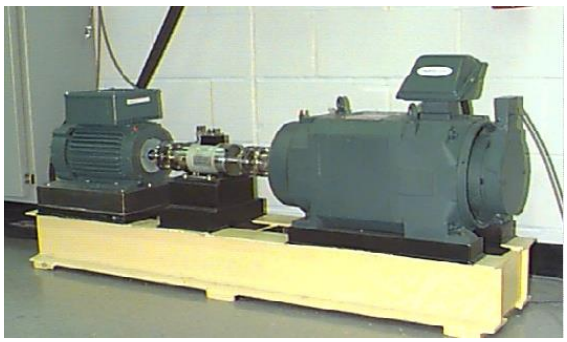


Figure 3. Experiment setup

3.2. Experimental results and analysis

Firstly, we extracted 8 time-domain features and 8 frequency-domain features respectively, and then a total of 16 features are obtained. In this study 200 data samples with 16 features from each fault type under each working condition were obtained, thus, the size of original target and source sample feature matrix is 1600*16. It should be pointed out that the training sample and testing sample belong to different working conditions.

To test the effectiveness of enhanced TCA for bearing fault diagnosis, a comparison study is performed between enhanced TCA and TCA_{share}, which only uses the shared topics to represent fault samples for classifier training. Furthermore, the enhanced TCA is also compared with principal component analysis (PCA), kernel principal analysis (KPCA), factor analysis (FA) and locally linear embedding (LLE) for feature selection. The mapped source data set are fed into SVM classifier for training, then the target data are used to test the trained model.

Two cases are studied in this paper: transfer among drive end bearings, and transfer between fan end bearing and drive end bearing. The classification results are shown in Tables 3 and 4, where the notation, for example D1797R1-D1797R2, is explained as follows: the left represents source data (for training only), the right is target data(used for testing), and 1797 represents the speed, R1, R2, R3, R4 stand for fault diameter with 0.1778mm, 0.3556mm, 0.5334mm, 0.7112mm respectively. The symbols F and D represent fan end bearing and drive end bearing respectively.

Table 3. Comparison among enhanced TCA and other algorithms for diagnosis accuracy under different working conditions using drive end bearing data

	Accuracy (%)					
	D1797R1 - D1797R2	D1797R1 - D1797R3	D1772R1 - D1775R1	D1772R1 - D1730R1	D1772R2 - D1730R2	Average
None	80.75	83.5	72.5	78	75.25	78
PCA	85.13	81	80	80	76.5	80.53
KPCA	81.5	81.5	77	79	75.63	78.93
LLE	77.5	79	82.13	80	79	79.53
FA	84	81.38	78	64	86.5	78.78
TCA_{share}	92	89	85	89	84.88	87.98
TCA	96.25	94.5	92.5	91	93.5	93.55

Table 4. Comparison among enhanced TCA and other algorithms for diagnosis accuracy using fan end bearing and drive end bearing data

	Accuracy (%)					
	F1797R1 - D1797R2	F1797R1 - D1797R3	F1772R1 - D1775R1	F1772R1 - D1730R1	F1797R2 - D1730R2	Average
None	71.63	84.5	71.5	78	71.5	75.43
PCA	86	80	81	81.75	79.88	81.73
KPCA	79.75	77.88	75.75	71	74	75.68
LLE	77.38	77	78	76.13	77.5	77.2
FA	80	80.5	82	74.13	76.5	78.63
TCA_{share}	90.5	85.63	85.5	90.25	80.5	86.48
TCA	92	95.75	96	90	92	93.15

The following conclusions can be obtained from Tables 3 and Table 4:

1) Using original features without any feature selection techniques as the input of SVM has the lowest performance.

2) The classification accuracy of traditional machine learning algorithms (PCA, KPCA, LLE and FA) comes to be only 78.88% on average, which means they are not suitable to train a precise diagnostic model for bearings under different conditions.

3) Compared with machine learning methods, the classification accuracy of transfer learning (TCA_{share}, enhanced TCA) is improved by about 8.31% and 14.47% respectively, and the best classification accuracy of enhanced TCA in both two cases is 96.25%. Thus, the enhanced TCA has the best classification performance.

4. Conclusion

A fault diagnosis method based on enhanced TCA strategy is proposed for bearing fault diagnosis under different working conditions. Here vibration data are viewed as a document consisted of a number of “fault word”, and with the K-means method, “fault word” can be generated. Compared with traditional machine learning algorithm and TCA_{share}, the enhanced TCA has the superior performance.

Acknowledgments

This project is supported by National Natural Science Foundation of China (Grant No. 51575102).

References

[1] Van M and Kang H J 2015 *J. IET Science Measurement Technology* **9** 671-680.

[2] Wu S D, Wu P H and Wu C W 2012 *J. Entropy* **14** 2650-54.

[3] Ndez-Francos D, Mart and Nez-Rego D 2013 *J. Computers & Industrial Engineering* **64** 357–365.

[4] Pandya D H, Upadhyay S H and Harsha S P 2013 *J . Expert Systems with Applications* **40** 4137–45.

[5] Inoue E, Ma R and Mitsuoka M 2011 *J . Ama Agricultural Mechanization in Asia Africa & Latin*

America **42** 89-93.
[6] Ali J B, Fnaiech N and Saidi L 2015 *J. Applied Acoustics* **89** 16-27.
[7] Sharma S, Abed W and Sutton R 2015 *J. IFAC-PapersOnLine* **48** 49-54.
[8] Yu Y and Junsheng C 2006 *J. Journal of Sound and Vibration* **294** 269-277.
[9] Shen Z, Chen X and Zhang X 2012 *J. Measurement* **45** 30-40.
[10] Shi Z, Song W and Taheri S 2016 *J. Entropy* **18** 70.
[11] Pan S J and Yang Q 2010 *J. IEEE Transactions on Knowledge and Data Engineering* **22** 1345-59.
[12] La L, Guo Q and Cao Q 2014 *J. Neural Computing and Applications* **24** 807-816.
[13] Pan W and Ming Z 2014 *J. IEEE Intelligent Systems* **29** 48-54.
[14] Hofmann T. 2001 *J. Machine Learning* **42** 177-196.
[15] Zhang Y, Brady M and Smith S 2001 *J. IEEE Transactions on Medical Imaging* **20** 45-57.
[16] Ahlgren P, Jarneving B and Rousseau R 2003 *J. Journal of the American Society for Information Science and Technology* **54** 550-560.