Expert Systems with Applications 38 (2011) 10199–10204

Contents lists available at ScienceDirect

Expert Systems with Applications

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

An effective procedure exploiting unlabeled data to build monitoring system

Xiukuan Zhao [a,b,*], Min Li [b], Jinwu Xu [b], Gangbing Song [c]

[a] Beijing National Observatory of Space Environment, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing 100029, China
[b] School of Mechanical Engineering, University of Science and Technology Beijing, Beijing 100083, China
[c] Department of Mechanical Engineering, University of Houston, Houston, TX 77204, USA

ARTICLE INFO

Keywords:
Active learning
Semi-supervised learning
Machine condition monitoring
Fault diagnosis
Support vector machine

ABSTRACT

Currently, condition-based maintenance becomes increasingly important with additions of factory automation through the development of new technologies. For many complicated machines, it is difficult to use mathematical models to describe their conditions. Intelligent maintenance makes it possible to perform maintenance similar to that of a human being. However, conventional artificial intelligent methods such as neural network and SVM use only labeled data (feature/label pairs) for training. Labeled instances are often difficult, expensive, or time consuming to obtain. Active learning and semi-supervised learning address this problem by using a large amount of unlabeled data together with labeled data to build better models. In this paper, a new active semi-supervised procedure was proposed to perform fault classification for machine condition monitoring. The effectiveness of the procedure was verified by its application to bearing diagnosis and gear fault detection.

1. Introduction

Since machine maintenance has significant impact in industry, it has received attention from researchers to practicing engineers. It is well known that maintenance cost is a major part of the total operating costs of all manufacturing and production plants. Intelligent maintenance systems make it possible to reduce maintenance costs. These systems perform maintenance routines as a human being would. Artificial intelligence (AI) enables the so-called intelligent maintenance system. Application of an expert system (ES) as a branch of AI in maintenance is one solution.

Support vector machine (SVM) is a relatively new computational learning method and can serve as an ES. It has been successfully applied to a number of applications, including face recognition, handwriting recognition, webpage classification, intrusion detection, and breast cancer diagnosis. It also has some applications in machine condition monitoring. In Fei and Zhang (2009), SVM with genetic algorithm was applied to fault diagnosis of a power transformer. Yuan and Chu (2006) presented a new multi-class SVM method. The effectiveness of the method was verified by its application to fault diagnosis of a turbo pump rotor. Abbasion, Rafsanjani, Farshidianfar, and Irani (2007) provided a procedure for fault classification of rolling bearings using an SVM classifier. Their method was applied to a three-phase induction motor test stand and achieved good performance. More applications of SVM in machine condition monitoring and fault diagnosis can be found in a review paper by Widodo and Yang (2007).

However, conventional SVMs use only labeled data for training. Labeled instances are often difficult, expensive, or time consuming to obtain, as they require the efforts of experienced human annotators. Meanwhile unlabeled data may be relatively easy to collect, though hard to use. In recent years, many methods that can be broadly divided into two groups, semi-supervised and active learning, have been proposed to solve such problems.

Semi-supervised learning algorithms are mainly based on three paradigms: density based methods (Chapelle, Sindhwani, & Keerthi, 2008; Joachims, 1999), graph-based algorithms (Belkin, Niyogi, & Sindhwani, 2006; Johnson & Tong, 2008; Niyogi, 2008), and boosting techniques (Saffari, Leistner, & Bischof, 2009). There are many applications using the semi-supervised learning method, such as robot control (Großmann, Wendt, & Wyatt, 2003), text classification (Ghani, 2002), diabetes diseases prediction (Wu, Diao, Li, Fang, & Ma, 2009) and image classification (Balcan et al., 2005). A survey on semi-supervised learning was presented in Zhu (2008).

The objective of active learning is to learn a function that accurately predicts the labels of new examples while requesting as few labels as possible. Pool-based active learning appears to be much more common among application papers. It has been studied for many real-world problem domains in machine learning, such as text classification (Hoi, Jin, & Lyu, 2006; Tong & Koller, 2002), image classification and retrieval (Tong & Chang, 2001; Zhang & Chen, 2002), video classification and retrieval (Yan, Jie, & Hauptmann,

* Corresponding author at: Beijing National Observatory of Space Environment, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing 100029, China.

E-mail address: kevin2008.zhao@gmail.com (X. Zhao).

0957-4174/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2011.02.078

2003), and cancer diagnosis (Liu, 2005). Settles (2009) presented a comprehensive survey about the literature of active learning.

Both semi-supervised learning and active learning can take advantage of the unlabeled data. It is quite natural to combine them to form a more effective method. Zhu, Lafferty, and Ghahramani (2003) proposed an approach to coupled active learning with semi-supervised learning using Gaussian fields and harmonic functions. Mao, Lee, Parikh, Chen, and Huang (2009) proposed a co-training framework, which is a kind of semi-supervised learning, to leverage unlabeled data to enhance intrusion detection and an integrated active labeling mechanism to extract an expert's knowledge only for uncertain instances. Yu, Varadarajan, Deng, and Acero (2010) proposed a unified global entropy reduction maximization (GERM) framework for active learning and semi-supervised learning for speech recognition.

In this paper, a new active semi-supervised method was proposed based on the pool query active learning and manifold regularization semi-supervised method, which is one of the graph based methods. The rest of the paper is organized as follows. Section 2 introduces the theoretical background of SVM, semi-supervised SVM, version space and active learning SVM. Section 3 then proposes our active semi-supervised method and presents a toy example. In Section 4 we present experimental results for rolling bearing fault diagnosis and gear fault detection. Finally, we offer our conclusions in Section 5.

2. Theoretical background

2.1. Support vector machine

Support vector machines (SVMs) are a set of related supervised learning methods used for classification and regression. Viewing input data as two sets of vectors in an n-dimensional space, SVMs construct a separating hyperplane in that space, which maximizes the margin between the two data sets.

Given training data form $\{\mathbf{x}_i, y_i\}$, where $i = 1, \ldots, l$, $y_i \in \{-1, 1\}$, $\mathbf{x}_i \in R^D$ ($\mathbf{x}_i$ has D attributes). The separating hyperplane can be described by $\mathbf{w} \cdot \mathbf{x} + b = 0$, where $\mathbf{w}$ is normal to the hyperplane and $|b|/\|\mathbf{w}\|$ is the perpendicular distance from the hyperplane to the origin. All the training data should satisfy the following constraints (Burges, 1998; Cortes & Vapnik, 1995):

$$y_i(\mathbf{x}_i \cdot \mathbf{w} + b) - 1 + \xi_i \geqslant 0 \quad \text{where} \quad \xi_i \geqslant 0 \ \forall i. \tag{1}$$

The positive slack variable ξ_i, $i = 1, \ldots, l$ is introduced to handle data that is not fully linearly separable. Subject to the constraints in (1), maximizing the SVM's margin is equivalent to finding

$$\min \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{l}\xi_i \quad \text{s.t.} \quad y_i(\mathbf{x}_i \cdot \mathbf{w} + b) - 1 + \xi_i \geqslant 0, \quad \xi_i \geqslant 0 \ \forall i, \tag{2}$$

where C controls the trade-off between the slack variable penalty and the size of the margin.

By solving the dual optimization problem of Eq. (2), the decision function is given by

$$f(\mathbf{x}) = sign\left(\sum_{i=1}^{l}\alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b\right), \tag{3}$$

where K is known as a Mercer kernel (Burges, 1998). The most popular non-linear kernels for classification are the radial basis kernel, the polynomial kernel and the sigmoid kernel. Using these non-linear kernels, the SVMs can achieve good performance when solving non-linear separable data.

2.2. Semi-supervised SVM

Semi-supervised learning methods use large amounts of unlabeled data together with labeled data to build better classifiers. Since semi-supervised learning requires less human effort and achieves higher accuracy, it is of great interest both in theory and in practice (Chapelle, Schölkopf, & Zien, 2006; Zhu, 2008). In this paper, we mainly focus on the manifold regularization semi-supervised method, which is one graph based method.

Graph-based semi-supervised methods define a graph where the nodes are labeled and unlabeled examples in the dataset and the edges (may be weighted) reflect the similarity of examples. These methods usually assume label smoothness over the graph. The manifold regularization framework is one kind of graph based methods. It employs two regularization terms (Belkin et al., 2006; Niyogi, 2008):

$$\min \frac{1}{l}\sum_{i=1}^{l} V(\mathbf{x}_i, y_i, f) + \gamma_A\|f\|_K^2 + \gamma_I\|f\|_I^2, \tag{4}$$

where V is an arbitrary loss function, K is a Mercer kernel, e.g. a linear or RBF kernel. I is a regularization term induced by the labeled and unlabeled data.

By using the manifold regularization framework, the SVM can be extended to Laplacian SVM (LapSVM) by solving the following problem (Belkin et al., 2006):

$$\min \frac{1}{l}\sum_{i=1}^{l}(1 - y_i f(\mathbf{x}_i))_+ + \gamma_A\|f\|_K^2 + \frac{\gamma_I}{(u+l)^2}\mathbf{f}^T L\mathbf{f}, \tag{5}$$

where $\mathbf{f} = [f(\mathbf{x}_1), \ldots, f(\mathbf{x}_{l+u})]^T$, and L is the graph Laplacian given by $L = D - W$. The diagonal matrix D is given by $D_{ii} = \sum_{j=1}^{l+u} W_{ij}$, and W_{ij} is edge weight in the data adjacency graph.

The standard SVM Eq. (2) is extended as:

$$\begin{aligned} &\min_{\alpha \in \Re^{l+u},\, \xi \in \Re^l} \quad \frac{1}{l}\sum_{i=1}^{l}\xi_i + \gamma_A \alpha^T K \alpha + \frac{\gamma_I}{(u+l)^2}\alpha^T KLK\alpha \\ &\text{s.t.} \qquad y_i\left(\sum_{j=1}^{l+u}\alpha_j K(\mathbf{x}_i, \mathbf{x}_j) + b\right) \geqslant 1 - \xi_i, \quad i = 1, \ldots, l, \\ &\qquad\qquad \xi_i \geqslant 0, \quad i = 1, \ldots, l. \end{aligned} \tag{6}$$

The dual form of Eq. (6) is formed as:

$$\begin{aligned} L_D = \max_{\beta \in \Re^l} \sum_{i=1}^{l}\beta_i - \frac{1}{2}\beta^T Q\beta \quad \text{s.t.} \quad 0 \leqslant \beta_i \leqslant \frac{1}{l}, \quad i = 1, \ldots, l \\ \text{and} \quad \sum_{i=1}^{l}\beta_i y_i = 0, \end{aligned} \tag{7}$$

where $Q = YJK(2\gamma_A I + 2\frac{\gamma_I}{(u+l)^2}LK)^{-1}J^T Y$.

LapSVM is implemented by using a standard SVM solver with the quadratic form induced by Eq. (7).

The decision function is:

$$f(\mathbf{x}) = \sum_{i=1}^{l+u}\alpha_i y_i K(\mathbf{x}_i, \mathbf{x}). \tag{8}$$

Experimental evidences and theoretical analyses suggest that the LapSVM algorithm is able to use unlabeled data effectively and obtain more accurate classifier (Belkin et al., 2006; Niyogi, 2008).

2.3. Version space and active learning SVM

A version space is the subset of all hypotheses that are consistent with the observed training examples (Mitchell, 1997). Active learning SVM can be well analyzed by using the version space. In the following, the concept of version space is briefly presented.

For the detail, please see (Herbrich, Graepel, & Williamson, 2006; Tong & Koller, 2002).

Given a set of labeled training data in the input space X, it can be mapped into a feature space F via a Mercer kernel. In the feature space, there is a set of hyperplanes that separate the data. This set of hyperplanes is called version space. In other words, hypothesis f is in version space if for every training instance $\mathbf{x}_i$ with label y_i we have that $f(\mathbf{x}_i) > 0$ if $y_i = 1$ and $f(\mathbf{x}_i) < 0$ if $y_i = -1$.

Let $\mathbf{w}$ be the unit vector of a hyperplane in F. The set of possible hypotheses is given as (Tong & Koller, 2002):

$$H = \left\{ f \middle| f(\mathbf{x}) = \frac{\mathbf{w} \cdot \Phi(\mathbf{x})}{\|\mathbf{w}\|}, \quad \mathbf{w} \in W \right\}, \tag{9}$$

where W denotes the parameter space and Φ is a kernel function. The version space V is defined as (Tong & Koller, 2002):

$$V = \{ f \in H | y_i f(\mathbf{x}_i) > 0, \quad i \in \{1, \ldots, n\}\}. \tag{10}$$

A duality can be found between unit vector $\mathbf{w}$ in W and hypotheses f in H. Thus the version space can be redefined as (Tong & Koller, 2002):

$$V = \{ \mathbf{w} \in W | \|\mathbf{w}\| = 1, \quad y_i(\mathbf{w} \cdot \Phi(\mathbf{x}_i)) > 0, \quad i \in \{1, \ldots, n\}\}. \tag{11}$$

Given a training sample $(\mathbf{x}_i, y_i)$, there is a hyperplane, $f(\mathbf{x})$ that can correctly classify it. According to the duality, the corresponding hyperplane in W bisects the surface of the hypersphere of $\|\mathbf{w}\| = 1$, and only the part satisfying $y_i(\mathbf{w} \cdot \Phi(\mathbf{x}_i)) > 0$ is favored. In this way, given a training set with n training samples, the version space is a connected region on the surface of the hypersphere, carved by the corresponding n hyperplanes. Fig. 1 illustrates such a version space in a three-dimensional parameter space carved by four hyperplanes.

As showed in Fig. 1, the surface of the hypersphere represents unit weight vectors. Each of the four hyperplanes corresponds to a labeled training instance. Each hyperplane restricts the area on the hypersphere in which consistent hypotheses can lie. Here, the convex polyhedron on top is a version space.

In a version space, an SVM classifier has the following geometrical interpretation. Imagining a largest hypersphere with its center restricted on the version space and not intersecting with any hyperplane, the normal vector of the optimal separating hyperplane $\mathbf{w}$ lies at the center of this largest hypersphere. The training samples are support vectors if they correspond to the hyperplanes which are tangent to this largest hypersphere (Tong & Koller, 2002; Wang, Chan, & Zhang, 2003).

In active learning SVM (A-SVM), $\mathbf{w}^* \in W$ is the unit parameter vector corresponding to the SVM that we would have known the actual labels of all of the data. We know that $\mathbf{w}^*$ must lie in each of the version spaces $V_1 \supset V_2 \cdots$, where V_i denotes the version space after ith query. Thus, we can reduce the space as fast as possible by shrinking the size of the version space as much as possible with each query. Assuming that only one unlabeled sample is selected in each learning cycle (Tong & Koller, 2002) proved that the hyperplane induced by this sample should halve the current version space, and proposed a "simple method" in which this sample is approximated by the unlabeled sample closest to the current separating hyperplane. The "simple method" is not the best method in identifying the desired sample; however, it is suitable for practical applications (Wang et al., 2003).

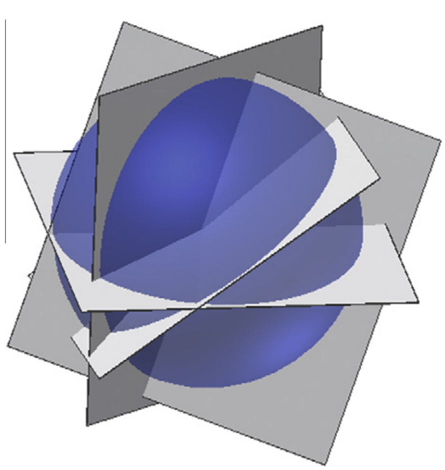

Fig. 1. A version space in a three-dimensional space.

3. Active semi-supervised learning

3.1. Theoretical analysis

Suppose we have a training set of labeled and unlabeled samples, and the number of labeled samples is too few for SVM to build a classifier with a reasonable level of performance. However, the unlabeled data can help to build a classifier with greater accuracy. The goal of active learning with semi-supervised SVM is to query the unlabeled samples that reveals the most information while taking into account the information already provided by the pool of unlabeled samples.

In Fig. 2, we have flattened out the surface of the unit weight vector hypersphere that appears in Fig. 1. The quadrilateral area is version space which is bounded by solid lines corresponding to labeled samples. As shown in Fig. 2, $\mathbf{w}_i$ corresponds to the ith SVM solution, $\mathbf{w}_i'$ is the ith LapSVM solution and $\mathbf{w}^*$ is the true solution that we would have known the actual labels of all of the data. The dotted line is the hyperplane induced by the candidate sample. According to the active learning SVM analyzed by Section 2.3, we should choose sample a to query next, as it is close to $\mathbf{w}_i$. However, as the LapSVM solution is more accurate than the SVM solution, $\mathbf{w}_i'$ should be closer to $\mathbf{w}^*$ than $\mathbf{w}_i$, or

$$\Delta \mathbf{w}_i' = \|\mathbf{w}_i' - \mathbf{w}^*\| < \Delta \mathbf{w}_i = \|\mathbf{w}_i - \mathbf{w}^*\|. \tag{12}$$

In the next iteration, $\Delta \mathbf{w}_{i+1}' < \Delta \mathbf{w}_{i+1}$. As shown in Fig. 2, it should be more effective to reduce the version space if we choose data close to $\mathbf{w}_i'$ than if we choose data close to $\mathbf{w}_i$. We call this active learning method with LapSVM solution Active LapSVM (A-LapSVM).

Given an unlabeled pool U, an A-LapSVM learner l has three components: $(f, q, X \cup U)$. The first component is a LapSVM classifier, as shown in (8), trained on the current set of the labeled data X and the unlabeled samples U. The second component q is the query function that decides which sample in U to query next. The A-LapSVM learner returns a classifier f after each query. The main difference between an active learner and a passive one is the query component q $(q = \arg\min_{x \in U} |f(x)|)$. This brings us to the issue of

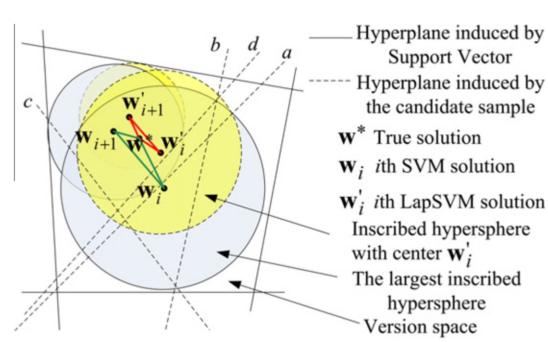

Fig. 2. The projection of the parameter space around the version space.

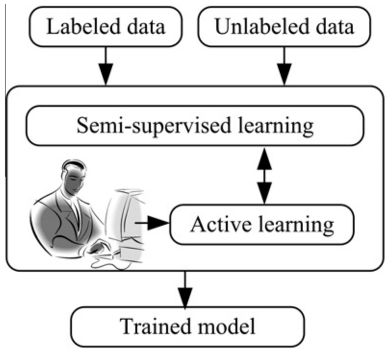

Fig. 3. The diagram of active semi-supervised learning.

choosing the next unlabeled sample to query. The difference between A-LapSVM and A-SVM is that A-LapSVM use the samples both in X and U to train f while A-SVM use only X to do the training.

The diagram of the active semi-supervised learning method is shown in Fig. 3. Labeled data and unlabeled data are used in semi-supervised learning (here, we choose LapSVM as the semi-supervised method). The active learning is processed based on the output of the semi-supervised learning; it will select the most informative data to let the oracle (e.g. a human expert) to give the label. The new labeled data is then used in the semi-supervised learning. The process iterates until the condition is satisfied and the output is then the trained model.

3.2. A toy example

We performed the A-LapSVM on the two moons data set (as shown in Fig. 4). The data set contains 200 examples with only 1 labeled example for each class at the very beginning. The red[1] diamond denotes positive type. The blue circle denotes negative type. The black points represent unlabeled examples and the red cross represents active selected example. Also shown are the decision surfaces of the A-LapSVM for each step. We chose RBF kernel as the kernel function and the kernel parameter was 0.35. The parameters in (6) were set to $\gamma_A = 0.2$, $\gamma_I = 0.5$.

Fig. 4 demonstrates how A-LapSVM success to find the most effective example to query next. We will achieve the optimal solution after five steps. The A-LapSVM decision boundary seems to be intuitively quite satisfying in step 6 (Fig. 4(f)). We only need to label seven examples other than label all of the 200 examples. In reality, it will reduce the labor cost of labeling to a great extend.

4. Experiment

4.1. Bearing fault diagnosis experiment

In order to verify the feasibility and the effectiveness of the proposed A-LapSVM method for machine fault diagnosis, signals of a rolling bearing rig at Case Western Reserve University were used (Loparo, 2003). Experiments were conducted using a 2-horsepower Reliance Electric motor and acceleration data were measured from the test rig.

Faults ranging from 0.007 to 0.040 in. in diameter were introduced separately at the inner raceway, rolling element (i.e. ball), and outer raceway. Faulty bearings were reinstalled into the test motor and vibration data were recorded for motor loads of 0 to 3-horsepower (motor speeds of 1797 to 1720 RPM). Ten conditions were considered in this study: normal, slight inner race fault, moderate inner race fault, severe inner race fault, slight outer race fault, moderate outer race fault, severe outer race fault, slight ball fault, moderate ball fault, and severe ball fault. Seven hundred and sixty instances were obtained for ten machine conditions (76 instances for each condition). One hundred and sixty instances were selected as the test samples (16 instances were randomly selected for each condition); the remaining instances were used as training samples.

Six time-domain features (root mean square, peak, skewness, kurtosis, shape factor, and impulse factor), four frequency-domain features (mean frequency, frequency center, root mean square frequency, and standard deviation frequency) and 32 wavelet coefficients energy (by five-layer wavelet packet decomposition) were respectively extracted to demonstrate the fault-related information.

Since these fault feature vectors had 42 dimensions and there existed correlations between them, the principal component analysis (PCA) (Bishop, 2006) was used for feature selection and dimension-reduction to decrease redundant variables in the data. By the PCA method, three principal components were chose to form the 3-dimensional fault feature vectors. These 3-dimensional fault feature vectors are shown in Fig. 5, in which the different shape graphics are data of different faults. It is visible that they are classifiable.

In the training phase, different numbers of training samples were used as labeled samples. SVM, A-SVM, LapSVM, and A-LapSVM methods were used to perform the classification of the machine conditions. During each test, it took 100 cycles to achieve the average accuracy. The RBF kernel was used as the mapping function. Genetic algorithm and cross-validation were used to optimize the parameters (RBF kernel parameter $\sigma = 0.0768$, $C = 0.6681$ in (2) and $\gamma_A = 0.7484$, $\gamma_I = 0.2606$ in (6)).

The accuracy rate of the classification based on different methods can be seen from Fig. 6. When the labeled data is very few, the

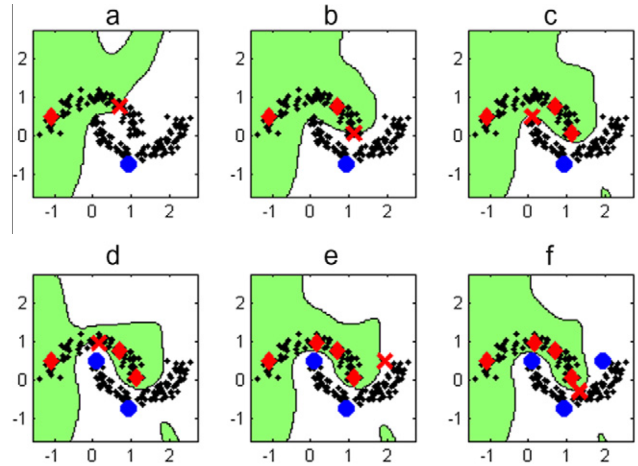

Fig. 4. Procedure of A-LapSVM with two moon data set.

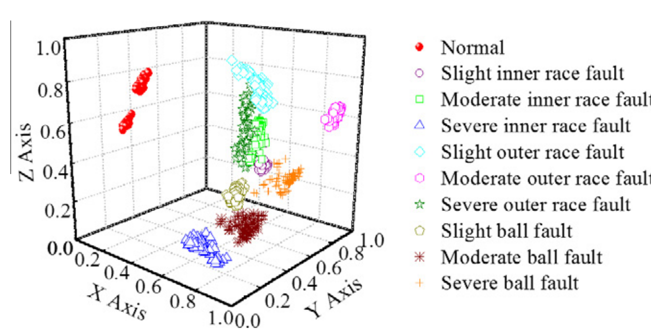

Fig. 5. Three-dimensional fault feature vectors.

[1] For interpretation of color in Fig. 4, the reader is referred to the web version of this article.

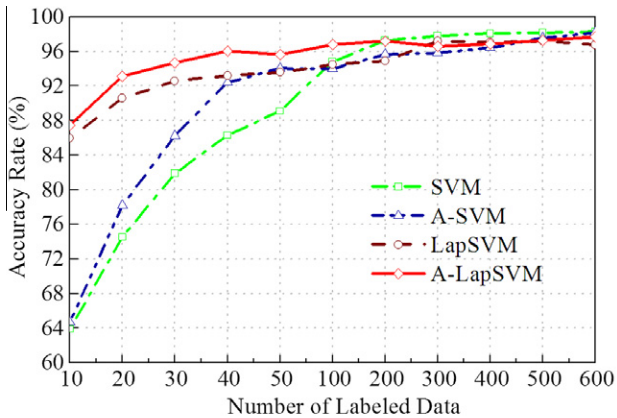

Fig. 6. The accuracy rate diagram of different methods.

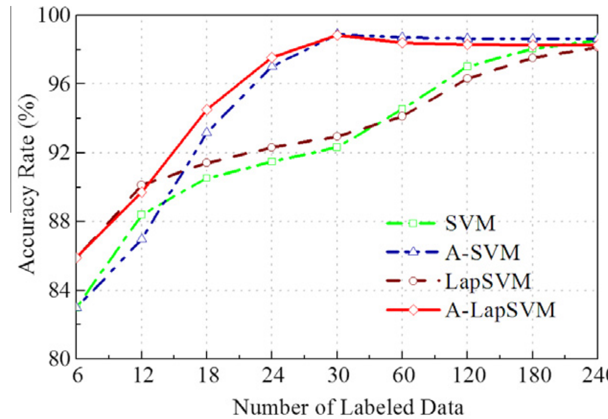

Fig. 8. The accuracy rate diagram of different methods.

A-LapSVM achieves better performance than other methods. The results verify the effectiveness of the proposed A-LapSVM method, especially when the number of labeled sample is very low.

The proposed A-LapSVM method only desires quit a few labeled data to achieve sufficient accuracy. As shown in Fig. 6, it requires only 40 labeled data to achieve 96% accuracy while other methods need about 200 labeled data to obtain the same accuracy. This method will select the most informative data instances for labeling by the users in the learning round; it is an efficiently interactive learning technique designed to reduce the labor cost of labeling. In other words, the proposed A-LapSVM method is to select an optimal set of unlabeled data instances that minimizes the expected risk of the learning round.

4.2. Gear fault detection experiment

Health monitoring of a gearbox experiment table with different gear fault was also used as an example to validate the proposed A-LapSVM method. Shown in Fig. 7 is the gearbox experiment table, which is installed in the Institute of Diagnosis and Monitoring at University of Science and Technology Beijing. The geared sleeve with different gear fault was used in the experiment. In this experiment, we chose six gear conditions: healthy, wearout fault, snaggletooth fault, backlash oversize, backlash undersize, and gear crack.

The speed of driving gear shaft was 900 RPM and the load produced by damper (shown in Fig. 7) was set to 5 Nm. B&K 4394 accelerometer, as shown in Fig. 7, was used to measure the vibration. 01dB data acquisition system was used to perform data record with a sampling rate of 16,384 Hz. A total of 1194 instances were obtained for the six gear conditions (199 instances for each condition). Two hundred and forty instances were selected as the training samples (40 instances were randomly selected for each condition); the remaining instances were used as test samples.

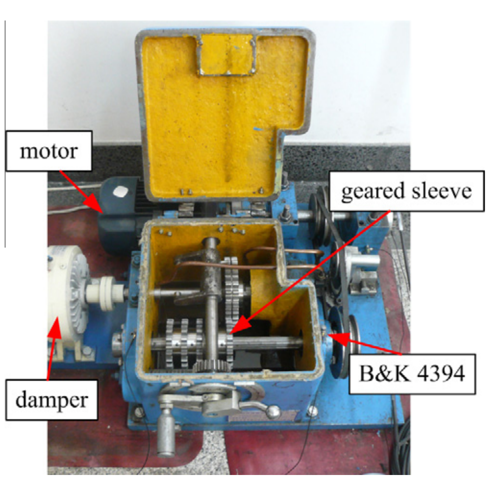

Fig. 7. Gearbox experiment table.

Four time-domain features (root mean square, peak value, skewness, and kurtosis) and six frequency-domain features (magnitude of axial frequency, axial frequency doubling, axial frequency tripling, rotation frequency, rotation frequency doubling, and rotation frequency tripling) were respectively extracted to reveal the fault-related information. PCA was also used for dimension-reduction and three principal components were chose to form the 3-dimensional fault feature vectors.

SVM, A-SVM, LapSVM, and A-LapSVM methods were used to perform the classification. In the training phase, different numbers of training samples were used as labeled samples. During each test, it took 100 cycles to achieve the average accuracy. The RBF kernel was used as the mapping function. Genetic algorithm and cross-validation were used to optimize the parameters (RBF kernel parameter $\sigma = 0.1443$, $C = 0.7703$ in (2) and $\gamma_A = 0.6491$, $\gamma_I = 0.3203$ in (6)).

The accuracy rate of the classification based on different methods can be seen from Fig. 8. When the labeled data is very few, the proposed A-LapSVM achieves better performance than other methods. The results indicate that active semi-supervised learning will be able to find the most informative data instance and significantly reduce the need for labeled instances in practice.

5. Conclusions

Machine condition monitoring is essentially a kind of pattern recognition or classification. We developed a new method A-LapSVM based on active learning and semi-supervised method with manifold regularization. The proposed method uses large amount of unlabeled data, together with the labeled data, to build better classifiers. The effectiveness of the method was analyzed under the version space theory framework and was also verified by its application to the bearing diagnosis and gear fault detection. Compared to the state-of-the-art SVM, A-SVM, and LapSVM, the proposed A-LapSVM achieves better performance when there are only a few labeled data. It is an efficiently learning technique designed to reduce the labor cost of labeling.

Acknowledgments

The reported research was supported by National Natural Science Foundation of China (No. 50705069), National Natural Science Foundation of China (No. 50674010), National High-tech R&D Program of China (No. 2007AA04Z169), and Research Fund for the Doctoral Program of Higher Education of China (No. 20070008050). Thanks are also due to the Case Western Reserve University, Rockwell Science Office of Naval Research and CVX, for the data sets of the rolling element bearing faults.

References

Abbasion, S., Rafsanjani, A., Farshidianfar, A., Irani, N., et al. (2007). Rolling element bearings multi-fault classification based on the wavelet denoising and support vector machine. *Mechanical Systems and Signal Processing, 21*(7), 2933–2945.

Balcan, M., Blum, A., Choi, P., Lafferty, J., Pantano, B., Rwebangira, M., (2005). Person identification in webcam images: An application of semi-supervised learning. In *The 22st ICML workshop on learning with partially classified training data, Bonn, Germany*.

Belkin, M., Niyogi, P., & Sindhwani, V. (2006). Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *Journal of Machine Learning Research, 7*(11), 2399–2434.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York: Springer.

Burges, C. J. C. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery, 2*(2), 121–167.

Chapelle, O., Schölkopf, B., & Zien, A. (2006). *Semi-supervised learning*. The MIT Press.

Chapelle, O., Sindhwani, V., & Keerthi, S. S. (2008). Optimization techniques for semi-supervised support vector machines. *Journal of Machine Learning Research, 9*, 203–233.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297.

Fei, S., & Zhang, X. (2009). Fault diagnosis of power transformer based on support vector machine with genetic algorithm. *Expert Systems with Applications, 36*(8), 11352–11357.

Ghani, R. (2002). Combining labeled and unlabeled data for multiclass text categorization. In *Proceedings of the nineteenth international conference on machine learning*. Morgan Kaufmann Publishers Inc..

Großmann, A., Wendt, M., & Wyatt, J. (2003). A semi-supervised method for learning the structure of robot environment interactions. *Cryptographic Hardware and Embedded Systems – CHES, 2003*, 36–47.

Herbrich, R., Graepel, T., & Williamson, R. (2006). *The structure of version space*. No. MSR-TR-2004-63, July 2004.

Hoi, S. C. H., Jin, R., & Lyu, M. R. (2006). Large-scale text categorization by batch mode active learning. In *Proceedings of the 15th international conference on world wide web*. Edinburgh, Scotland: ACM.

Joachims, T. (1999). Transductive inference for text classification using support vector machines. In *Proceedings of the sixteenth international conference on machine learning*. Morgan Kaufmann Publishers Inc..

Johnson, R., & Tong, Z. (2008). Graph-based semi-supervised learning and spectral kernel design. *IEEE Transactions on Information Theory, 54*(1), 275–288.

Liu, Y. (2005). Active learning with support vector machine applied to gene expression data for cancer classification. *Journal of Chemical Information and Computer Sciences, 36*(6).

Loparo, K. A. (2003). *Bearings vibration data set*. Case Western Reserve University, Sep. 20, 2009.

Mao, C., Lee, H., Parikh, D., Chen, T., Huang, S., et al. (2009). Semi-supervised co-training and active learning based approach for multi-view intrusion detection. In *Proceedings of the 2009 ACM symposium on applied computing*. Honolulu, Hawaii: ACM.

Mitchell, T. (1997). *Machine learning*. McGraw-Hill.

Niyogi, P. (2008). *Manifold regularization and semi-supervised learning: Some theoretical analyses*. University of Chicago.

Saffari, A., Leistner, C., & Bischof, H. (2009). Regularized multi-class semi-supervised boosting. In *Proceedings of IEEE society conference on computer vision and pattern recognition, Miami Beach, FL, USA*.

Settles, B. (2009). *Active learning literature survey*. Madison: University of Wisconsin.

Tong, S., & Chang, E. (2001). Support vector machine active learning for image retrieval. In *Proceedings of the ninth ACM international conference on multimedia*. Ottawa, Canada: ACM.

Tong, S., & Koller, D. (2002). Support vector machine active learning with applications to text classification. *Journal of Machine Learning Research, 2*, 45–66.

Wang, L., Chan, K. L., & Zhang, Z. (2003). Bootstrapping SVM active learning by incorporating unlabelled images for image retrieval. In *IEEE computer society conference on computer vision and pattern recognition* (Vol. 1, pp. 629–634).

Widodo, A., & Yang, B.-S. (2007). Support vector machine in machine condition monitoring and fault diagnosis. *Mechanical Systems and Signal Processing, 21*(6), 2560–2574.

Wu, J., Diao, Y., Li, M., Fang, Y., Ma, D., et al. (2009). A semi-supervised learning based method: Laplacian support vector machine used in diabetes disease diagnosis. *Interdisciplinary Sciences: Computational Life Sciences, 1*(2), 151–155.

Yan, R., Jie, Y., & Hauptmann, A. (2003). Automatically labeling video data using multi-class active learning. *Ninth international conference on computer vision* (Vol. 1, pp. 516–523). France: Nice.

Yu, D., Varadarajan, B., Deng, L., & Acero, A. (2010). Active learning and semi-supervised learning for speech recognition: A unified framework using the global entropy reduction maximization criterion. *Computer Speech & Language, 24*(3), 433–444.

Yuan, S., & Chu, F. (2006). Support vector machines-based fault diagnosis for turbo-pump rotor. *Mechanical Systems and Signal Processing, 20*(4), 939–952.

Zhang, C., & Chen, T. (2002). An active learning framework for content-based information retrieval. *IEEE Transactions on Multimedia, 4*(2), 260–268.

Zhu, X., Lafferty, J., & Ghahramani, Z. (2003). Combining active learning and semi-supervised learning using Gaussian fields and harmonic functions. In *ICML 2003 workshop on the continuum from labeled to unlabeled data in machine learning and data mining* (pp. 58–65).

Zhu, X. (2008). *Semi-supervised learning literature survey*. University of Wisconsin-Madison, Computer Science.