21st International Conference on Pattern Recognition (ICPR 2012)
November 11-15, 2012. Tsukuba, Japan

Semi-Supervised Adaptive Parzen Gentleboost Algorithm for Fault Diagnosis

Chengliang Li, Zhongsheng Wang, Shuhui Bu, Zhenbao Liu
School of Aeronautics, Northwestern Polytechnical University, Xi'an 710072
licous@mail.nwpu.edu.cn

Abstract

In this paper, we present a novel semi-supervised strategy for machine fault diagnosis. In the proposed method, we select parzen window as the generative classifier and Gentleboost as the discriminative classifier. Compared with SVM, boosting method has a very interesting property of relative immunity to overfitting. In addition, we propose a novel adaptive parzen window algorithm. It employs variational adaptive parzen window rather than a global optimized and fixed window, therefore, more accurate density estimates can be obtained. In experiments, artificial and machine vibration data are used to compare with other algorithms. Our proposed algorithm achieves stronger robustness and lower classification error rate.
.

1. Introduction

Recently, Neural network and SVM have been widely applied in the field of intelligent fault diagnosis [2, 3]. Traditional pattern recognition methods such as supervised learning methods utilize labeled samples for training. However, selecting the labeled samples is a time consuming task.

The methods using both labeled and unlabeled data to train classification model are called semi-supervised learning [6, 9, 11]. Self-training is a classical semi-supervised learning method [12] . A classifier is firstly trained with the labeled data, and then the classification model is used to classify the unlabeled data. Thus, the unlabeled samples, which are classified with most confident scores, are incrementally added to the training set with their predicted labels until convergence is reached. However, this semi-supervised training algorithm has a limitation that it does not give good results only using discriminative classifiers. Because the samples could be classified in the wrong class with high confident scores, and consequently led to low classification accuracy. In order to overcome the problem, a Help-training method [1] has been proposed. In the method, a generative model, which does not focus only on the boundary between classes, is adopted to help discriminative classifier makes better decisions during the self-labeling process.

In this study, we propose a novel Help-train algorithm named Semi-Supervised Adaptive Parzen Gentleboost (SSAPG), which can achieve more robust and accurate classification results. In the proposed method, an adaptive parzen window method which can estimate accurate density estimates is used as generative model, and a Gentleboost method is used as discriminative classifier which can immunity to overfitting [5]. From the simulation and experiments data, we can conclude that proposed method outperform conventional methods.

The rest of this paper is organized as follows. In section 2, we describe Semi-Supervised Adaptive Parzen Gentleboost algorithm (SSAPG). The experimental results on both artificial and machine vibration data are given in section 3. We conclude our work in section 4.

2. Semi-Supervised Adaptive Parzen Gentleboost algorithm

We briefly describe Help-Train algorithm [1]. Let us consider the main classifier C which is based on a discriminative approach, and the classifier G based on a generative model. The classifier G produces a probability density model, and it is used to select samples which have a high probability to belong to a class. These selected samples constitute the candidate samples for labeling process. The classifier C then classifies the pre-selected samples, and those that are classified with most confident scores are added to the training set. The process is repeated until all unlabeled data are labeled. The detailed description of classifiers C and G are introduced in the next sub-sections.

2.1.Supervised Gentleboost (SG)

Comparing with SVM, boosting method only requires adjusting the number of cycles to control the training accuracy. In addition, the most interesting property of boosting is its relative immunity to overfitting [5]. In this study, we adopt Gentleboost algorithm as the main classifier which has better performance and faster data detection than Adaboost algorithm [4]. Like SVM [1], the magnitude of the margin can be interpreted as a measure of confidence about the decision of the classifier with respect to a sample. In Adaboost, the margin of a training example with respect to a classifier is defined as:

$$margin_f(\boldsymbol{x},y)=yF(\boldsymbol{x})/\sum_{m=1}^{M}\alpha_m \tag{1}$$

In Gentleboost, the parameter $\alpha_m \equiv 1$. The margin lies in the interval [-1, 1] and is positive if and only if the respective pattern is classified correctly [5, 7].

2.2. Adaptive parzen window

In this section, we present the proposed adaptive method which can improve the non-parametric technique performance in probability density estimation by the sparseness degree relationship, It is applied into a Help-training framework. The probability densities of two classes are defined as:

$$G_+ = \frac{1}{l_+}\sum_{i|y_i=+1} k_p(\boldsymbol{x}_i,\boldsymbol{x})$$
$$G_- = \frac{1}{l_-}\sum_{i|y_i=-1} k_p(\boldsymbol{x}_i,\boldsymbol{x}) \tag{2}$$

where l is the number of samples, and k_p is the parzen kernel function. we choose Gaussian kernel function as follows:

$$k_p(\boldsymbol{x}_i,\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{m}{2}} h^m} exp\text{-}(\frac{(\boldsymbol{x}-\boldsymbol{x}_i)^T(\boldsymbol{x}-\boldsymbol{x}_i)}{2h^2}) \tag{3}$$

where h is the bandwidth, and m is the dimension.

We estimate the sparseness degree around a sample based on its distance to its neighbors. The bandwidth adaptation method is as follows: First, a global optimal fixed bandwidth h is decided [8]. Second, adjust the variable bandwidth h_i around h and let them be proportional to the sparseness degrees of their nearby regions. N_i denotes the set of neighbors of $\boldsymbol{x}_i$, k is the number of classes.

First, the mean square of local bandwidth h_i is equal to square of global optimized bandwidth h.

$$\frac{1}{l}\sum_{i=1}^{l} h_i^2 = h^2 \tag{4}$$

For two samples $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ in the same set, according to the sparseness degree relationship, the local bandwidths are defined as:

$$h_i^2 = c_1 \sum_{\boldsymbol{x}_k \in \mathrm{N}_i} \|\boldsymbol{x}_k - \boldsymbol{x}_i\|^2 \tag{5}$$

$$h_j^2 = c_2 \sum_{\boldsymbol{x}_k \in \mathrm{N}_j} \|\boldsymbol{x}_k - \boldsymbol{x}_j\|^2 \tag{6}$$

where c_1 and c_2 are constant weight coefficients. If the c_1=c_2, we can compute the proportion relationship between h_i and h_j.

$$\frac{h_i^2}{h_j^2} = \frac{\sum_{\boldsymbol{x}_k \in \mathrm{N}_i} \|\boldsymbol{x}_k - \boldsymbol{x}_i\|^2}{\sum_{\boldsymbol{x}_k \in \mathrm{N}_j} \|\boldsymbol{x}_k - \boldsymbol{x}_j\|^2} \tag{7}$$

In order to obtain the variable bandwidth h_i, the numerator term is kept unchanged and let the denominator accumulate in eq. (7). And then, adopting the eq.(4), the variable bandwidth h_i can be derived as:

$$h_i = (lh^2 \frac{\sum_{\boldsymbol{x}_k\in \mathrm{N}_i}\|\boldsymbol{x}_k-\boldsymbol{x}_i\|^2}{\sum_{i=1}^{l}\sum_{\boldsymbol{x}_k\in \mathrm{N}_i}\|\boldsymbol{x}_k-\boldsymbol{x}_i\|^2})^{1/2} \tag{8}$$

Finally, the adaptive Gaussian kernel can be computed as follows:

$$k_p(\boldsymbol{x}_i,\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{m}{2}} h_i^m} exp\text{-}(\frac{(\boldsymbol{x}\text{-}\boldsymbol{x}_i)^T(\boldsymbol{x}\text{-}\boldsymbol{x}_i)}{2h_i^2}) \tag{9}$$

2.3. Proposed algorithm

The proposed algorithm (SSAPG) utilizes the supervised Gentleboost as discriminative classifier and adaptive parzen window as generative model. The flow of the proposed algorithm is provided in Table 1. In the algorithm, $\mathbf{a}_0$ is the initial threshold for selecting most confident unlabeled samples.

Table 1. Proposed algorithm (SSAPG)

Input **L**=labeled samples, **U**=unlabeled samples
Output Parameters of **Gentleboost**

-**Initialize** the working set **W**=**L** and **a**=$\mathbf{a}_0$
While **U**≠Φ do
- -Train **Gentleboost** with the working set **W**
- -Estimate the probability density model **adaptive parzen+** for positive samples in **W** and estimate the probability density model **adaptive parzen-** for negative samples in **W**
- -Select $\mathbf{n}_1$ samples from **U** with high probability according to **adaptive parzen+** and select $\mathbf{n}_2$ samples from **U** with high probability according to **adaptive parzen-**
- -Compute the output of the **gentleboost** for the selected **($\mathbf{n}_1$+$\mathbf{n}_2$)** samples
- -Constitute the set S formed by the samples whose output are most confident, ***f*(*x*)**>=**a**
- -Update the working set **W**←**W** ∪ **S**
- -Update the unlabeled set **U**←**U-S**
- -Reduce the value of **a** if **S**=Φ

End while
-Return the final parameters of the **Gentleboost**

3. Experiments

In order to evaluate the performance of the proposed methods, we performed experiments for two different classification tasks, including: (1) Two moons data problem, (2) CWRU bearing test data [13].

3.1. Performance evaluation

In this section, the proposed method was tested on the Two Moons data set as shown in Figure. 1, which is a standard benchmark for the semi-supervised learning algorithms used in the literature [1]. In each class, we randomly selected one point to form the labeled data and the remaining data served for unlabeled samples. The task is to predict the class for the unlabeled samples. In this experiment, we compared our algorithm with the results of ∇ S^3VM (deltaS^3VM), cS^3VM, DA, TSVM/SVMlight, Self-Training SVM (STSVM) [10] , Help-Train SVM (HTSVM) [1] in same condition. For this synthetic problem we used Gaussian kernel with same hyper-parameters h=0.5 same as in [10], and the a_0=0.9.

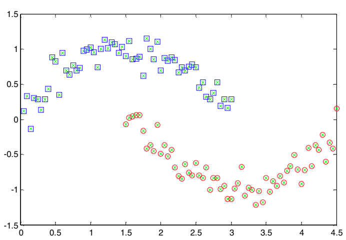

Figure 1. Illustration of two moons problem

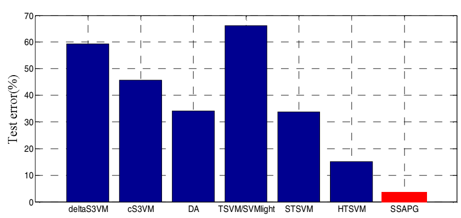

Figure 2.Test error rate on the Two Moons dataset

In Figure 2, it appears clearly that SSAPG performs better than STSVM and HTSVM. In addition, we note that our method has lower test error compared with other algorithms.

3.2. Fault diagnosis

In this experiment, we adopted the experiment data provided by CWRU bearing test data center [13]. The ball bearings used in the experiment were installed in a motor driven mechanical system. Single point faults with a diameter of 0.007 inches were introduced separately at inner race, ball, and outer race of the drive-end bearings using electro-discharge machining. The motor has a fixed speed of 1772 rpm is employed during the experiment. The bearing vibration signals of four bearing conditions were captured, including normal condition and three fault conditions with a sampling frequency of 12kHz. In each condition, 228 samples were captured, and five statistical characteristics were extracted (mean, standard, variance, skewness, kurtosis). Each class takes 28 labeled samples and 100 unlabeled samples as training set. Remaining 100 samples compose a test set in each class.

Firstly, the proposed method is compared with the SG algorithm in the training accuracy using the training set. Figures 3-5 present the training error rate with the number of repeat for outer, inner and balling fault classification.

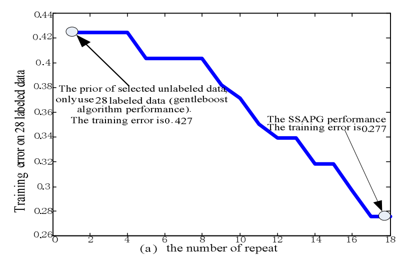

Figure 3. Outer fault training error rate under the number of repeat

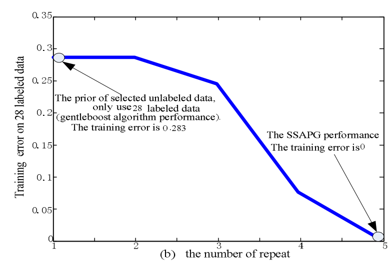

Figure 4. Inner fault training error rate under the number of repeat

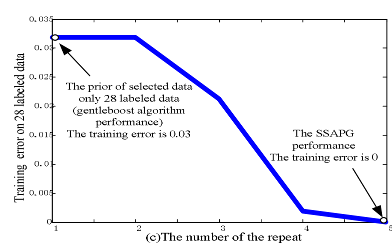

Figure 5. Ball fault training error rate under the number of repeat

Secondly, the proposed algorithm is compared with the SG algorithm in test accuracy with the increase the different ratio of labeled data. In Figures 6-8, each presents the classification error with different number of labeled samples. Figure 6 depicts the results of outer fault classification, Figure 7 depicts the inner fault classification, and Figure 8 depicts the balling fault classification.

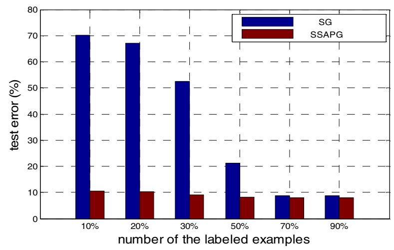

Figure 6. Test error of SSAPG algorithms with respect to the outer fault with different number of labeled samples.

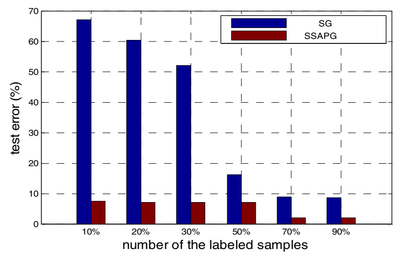

Figure 7. Test error of SSAPG algorithms with respect to the inner fault of number of labeled samples

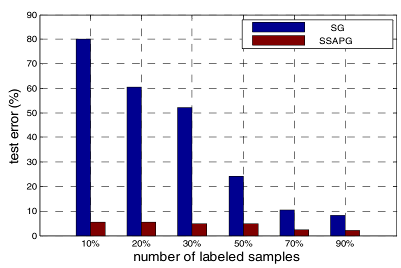

Figure 8. Test error of SSAPG algorithms with respect to the balling fault of number of labeled samples

From the results, we conclude that the test error rates obtained from the proposed method are smaller, and it less depended on the number of labeled samples in the training set. Thus, the proposed method is superior, and also makes the resulting classifier more robust.

4. Conclusion

A novel semi-supervised pattern recognition method (SSAPG) is proposed in this research. An adaptive parzen window is proposed for achieving better density estimates. In addition, we select Gentleboost as the discriminative classifier to overcome the overfitting problem. We evaluate the proposed algorithm on artificial data, and the results indicated that our algorithm performs better than other methods. Our method is also applied to fault diagnosis and compared with SG, the experiment results indicate that proposed method achieves better accuracy and robustness.

In the future work, we plan to use other generative models instead of parzen windows in order to improve the efficiency during the training process.

Acknowledgments

This work was supported in part by the Natural Science Foundation of China (NSFC) (51075330, 61003137, 50975231) and the Scientific Research Foundation (SRF) for Returned Overseas Chinese Scholars (ROCS), Northwestern Polytechnical University (NWPU) fundamental fund, and Natural Science Basic Research Plan in Shanxi Province of China (2012JQ8037).

References

[1] M.M. Adankon, M. Cheriet., Help-Training for semi-supervised discriminative classifier. Application to svm, *19th International Conference on Pattern Recognition*, 2008.

[2] Li B., Chow M., Hung J., Neural-Network-Based Motor Rolling Bearing Fault Diagnosis. *IEEE Transactions On Industrial Electronics*, 2000.

[3] Widodo, A, Yang, B-S., Support vector machine in machine condition monitoring and fault diagnosis. *Mechanical Systems and Signal Processing,* 2007

[4] W Lior, M Ian ., Robust boosting for learning from few examples, *IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, 2005.

[5] Richard O., Duda Peter E., Hart David G.Stork, *Pattern Classification. fourth Edition* , 2009.

[6] M. Seeger, Learning with labeled and unlabeled data, *Technical Report*, Institute for Adaptive and neural Computation, University of Edinburgh, 2001.

[7] B. E. Boser, I. Guyon, and V. Vapnik. A training algorithm for optimal margin classifiers. *In Computational Learing Theory*, 1992.

[8] Hallp, Kang.K., Bandwidth choice for nonparametric classification, *The Annals of statistic*, 2005.

[9] M. Belkin, P. Niyogi, and V. Sindhwani. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *Journal of Machine Learning Research*, 2006.

[10] V. S. Chapelle, O. and S. S. Keerthi. Branch and bound for semi-supervised support vector machines. In *Proceedings of the NIPS* , 2006.

[11] Zhu X., Semi-supervised learning literature survey. *Technical Report 1530*, Computer Sciences, University of Wisconsin-Madison, 2007.

[12] D. Yarowsky. Unsupervised word sense disambiguation rivaling supervised methods. *In Meeting of the Association for Computational Linguistics*, 1995.

[13] Loparo KA Bearings vibration data set, CWRU, http://www.eecs.case.edu/laboratory/bearing/.