ATLANTIS PRESS
Advances in Computer Science Research, volume 63
International Conference on Artificial Intelligence and Engineering Applications (AIEA 2016)

An Effective and Novel Weighted Support Vector Machine Method for Control Chart Pattern Recognition

Jianping Chen [a], Beixin Xia [b, *] and Xin Chen [c]

School of Mechatronic Engineering and Automation, Shanghai University, Shanghai, 200072, China

[a]jackiechan@shu.edu.cn, [b]bxxia@shu.edu.cn, [c]zgtcx@163.com

Abstract. Control chart pattern recognition is the method to realize quality online monitoring and diagnosis of production process. For the conditions that the number of existing normal mode products is much higher than the abnormal ones during the actual manufacturing process, we proposed a method about WSVM (Weighted Support Vector Machines) for dynamic process of abnormal pattern recognition based on PCA (Principal Component Analysis). We put the proposed method into our experiment, the experimental simulation results show that this method proposed in this paper has a big advantage over the existing SVM (Support Vector Machine) on highly imbalanced classification problem, which suitable for quality monitoring and diagnosis of dynamic production process.

Keywords: Control Chart Pattern Recognition; WSVM; Quality Monitoring and Diagnosis; Imbalanced Classification Problem.

1. Introduction

Control chart is the important tool for monitoring and diagnostic of dynamic production process . In recent years, the scholars found that the traditional algorithm can be used only when the number of samples in each class is approximately of the same size (balanced problems). When this is not the case, the traditional algorithm need to be optimized in order to address the imbalanced classification problem aspect.

However, there is few CCPR algorithm that takes into consideration the imbalanced nature of the abnormal pattern detection problems. If we use traditional generic SVM to solve the highly imbalanced classification problem, we will find it have poor performance. In the present study, we proposed a PCA-WSVM approach for the special abnormal quality control detection problems. First of all ,we should pre-process the raw data generated by the dynamic machining process , the method of PCA can not only reduce the dimensions of the original data effective but also highlight the main features of information , remove the noise of the raw data . Finally, put the processed data into the WSVM classifier to train them. We compared PCA-MSVM with SVM on imbalanced classification problem with respect to G-mean. The experimental results show that: we can see that PCA-WSVM demonstrates a more robust behavior.

2. Simulation Experiment

2.1 Pattern Generation.

The mathematical model for all components considered in this study can be written as:

$$x(t) = D + d(t) + r(t) \tag{1}$$

We can found from the above formula: These simulated control charts ($x(t)$) consist of three major components: $x(t)$ represents the target of the dynamic production process the target , D represents the average quality parameters under controlled conditions , $r(t)$ represents the fluctuation error caused by random factors , $d(t)$ represent fluctuating error due to abnormal factors .

By Referring to previous works [1, 2], in order to simplify the simulation data , D will be set to zero , $r(t) \sim N(0\,, \sigma^2)$.

2.2 Performance Measures

In this study, we employed the measures for evaluating the performance of proposed algorithms: data mining based measures especially tailored for imbalanced problems.

Different performance measures can be used to used for imbalanced data classifications. Such measures can be obtained, directly or indirectly, from the classification confusion matrix (table1). For a classification problem with k classes , the confusion matrix is a square matrix $\in R^{k\times k}$, with each of its matrix c_{ij} , denoting the percentage of the samples that belong to the class i and classified to the class j . For the special case of binary classification (positive and negative) , the confusion matrix is as follows: where *TP* , *FP* , *FN* , *TN* stand for true positives , false positives , false negatives and true negatives correspondingly . Obviously, the higher the values of the main diagonal the better the classifier is . A common performance measure for classification is the so-called accuracy, which is expressed as the correctly classified samples over the total number of training samples. However, for imbalanced classification problems this might not be a good performance indicator, since the majority class dominates the behavior of this metric more specifically, native decision rules can yield high classification accuracy. Alternatively, sensitivity and specificity can be used. They are defined as

$$Sensitivity = \frac{TP}{TP + FN}, Specificity = \frac{TN}{TN + FP}$$

Table.1 Confusion matrix for binary classification problem

	Positive class	Negative class
Positive class	*TP*	*FP*
Negative class	*FN*	*TN*

Specificity is not and therefore, it is a more appropriate measure for this purpose. Another metric often used is the geometric mean of sensitivity and specificity (often abbreviated *G*-mean) which is defined as the square root of the product between sensitivity and specificity. In this study, sensitivity, specificity and *G*-mean are used as performance measures.

3. Results and discussion

In this paper, we compared PCA-MSVM against SVM for imbalanced classification problem. We use sensitivity, specificity and G-mean as performance measures.

Experiments on both SVM and WSVM were conducted with Libsvm, which was developed by professor Lin [3]. For each classification problem, we generate a total of 1000 data points and for cross validation purposes, 90% of the data was used for training and the rest 10% was used for testing.

To the best of author's knowledge, there are few kernel functions are used to solve these imbalanced nature of the abnormal pattern detection problems, so we chose the Radial Basis Function (RBF) kernel. For this kernel function, the similarity between two data points x_i and x_j are given by:

$$k(x_i, x_j) = exp(-\gamma\|x_i - x_j\|^2), \gamma \geq 0 \quad (2)$$

For each class, the weights are estimated as the inverse of the class size:

$$C^+ = \frac{C}{n^+}, C^- = \frac{C}{n^-} \quad (3)$$

By referring to previous works [4-7], where n^+ and n^- are the size of normal and abnormal class, $C^+$and C^- are weights corresponding to the normal and abnormal classes respectively. Note that for balanced problems the weights become equal ($C^+ = C^-$) and the algorithm reduces to SVM.

This experiment we will generate 1000 data, We consider the size of imbalanced normal and abnormal data as $(50 + r)\%$ and $(50 - r)\%$ respectively. The two classification schemes (SVM and PCA-SVM) were compared terms of their G-mean for different values of radio *r*. We provide the table that shows the result（see table2）.

Table 2 Sensitivity (Sen), Specificity (Spe), Accuracy (Acc), and G-means (G) of SVM and PCA-WSVM for different imbalanced radio

r	SVM			PCA-WSVM		
	Sen	*Spe*	*G*	*Sen*	*Spe*	*G*
5	0.88	0.84	0.86	0.89	0.88	0.89
10	0.90	0.82	0.86	0.88	0.88	0.88
15	0.91	0.80	0.86	0.89	0.86	0.88
20	0.94	0.77	0.85	0.86	0.89	0.87
25	0.94	0.73	0.83	0.88	0.88	0.88
30	0.94	0.72	0.82	0.89	0.87	0.88
35	0.96	0.69	0.81	0.91	0.84	0.87
40	0.99	0.60	0.76	0.84	0.91	0.88
45	0.99	0.51	0.70	0.93	0.93	0.93

From the table, we can find that PCA-WSVM demonstrates a more robust behavior for various values of *r*. We must note that in many cases SVM achieves high sensitivity and very low specificity. Low classification tells us that all the test samples are classified in the same class and is indicative of poor performance.

4. Conclusion

For highly imbalanced binary classification, we proposed a PCA-WSVM approach which is more robust than traditional SVM. We test the two algorithms for binary classification in imbalanced environment. The result demonstrates that PCA-WSVM is much better than SVM in terms of specificity and *G*-mean.

Much work we need to do, we must spend more time to study multi-class classification. Since in reality one is interested to discriminate not only the normal versus abnormal problem but have as much information as possible about the abnormality , multi-class CCPR need to receive more attention in future works .

References

[1] C. Cheng, H. Cheng, K. Huang. A support vector machine-based pattern recognizer using selected features for control chart patterns analysis. International Conference on Industrial Engineering and Engineering Management.2009, p.419–423.

[2] Yang, S.-F. Process control using VSI because selecting control charts. Journal of Intelligent Manufacturing. Vol.21(2009)No.6, p.853 – 867.

[3] Gandomkar M, Vakilian M. Ehsan M, A Genetic-based Tabu Search Algorithm for Optimal DG Allocation in Distribution Networks. Electricty Power Component System. Vol.33 (2005) No.12, p.13511-1361.

[4] S. Liu, C.-Y. Jia，H. Ma, A new weighted support vector machine with GA-based parameter selection. Proceedings of 2005 international conference machine learning and cybernetics. 2005, p. 4351-4355.

[5] Du and Chen, S. Du, S. Chen. Weighted support vector machine for classification. IEEE international conference systems. 2005, p. 3866- 3871.

[6] Y.-M. Huang, S.-X. Du. Weighted support vector machine for classification with uneven training class sizes. Proceedings of 2005 international conference on machine learning and cybernetics. 2005, p. 4365 – 4369.

[7] J. P. Hwang, S. Park, E. Kim, A new weighted approach to imbalanced data classification problem via support vector machine with quadratic cost function. Expert Systems with Applications. Vol.38(2011)No.7, p.8580 – 8585.