Hyperparameter Selection with Good Region Recognition for SVM Based Fault Diagnosis

Siyu Lei, Yue Guo
Nanchong Vocational and Technical College, Sichuan, Nanchong, 637131

Abstract: This paper proposes a novel method of good region recognition for hyperparameter selection of SVM. The method can provide a much smaller good region for optimization search-based methods, and thus it can greatly save computation time. Experimental results show that the proposed method improves efficiency of fault diagnosis of rolling bearing with no accuracy loss.

Keywords: good region recognition; hyperparameter selection; support vector machine; fault diagnosis

1.Introduction

Support vector machine (SVM) was firstly proposed by Vapnik and has been widely used in data mining tasks recently, because it has a strong theoretical background. SVM applies the structure risk minimization principle that has been proven to be superior over the empirical risk minimization principle .

As a good data classifier, SVM can classify input data with different labels by finding a set of support vectors. In the study of improving learning and generalization ability, a crucial step is to optimize hyperparameters for SVM. The hyperparameters include a penalty coefficient C in the basic theory of SVM and the kernel parameter in the kernel. At present, the Gaussian kernel is the most common kernel function due to its many good properties and σ is the only kernel parameter in the Gaussian kernel.

The most frequently-used method to get the best hyperparameter combination is grid search . Although it has advantages of stability and intuitive idea, large computation time is not a negligible disadvantage of grid search.

This paper aims to speed up the selection process by reducing the selection region of hyperparameters with a model of good region recognition. A novel method based on the idea of good region recognition for hyperparameter selection is introduced in Section 2. Section 3 presents experimental results of the proposed method to optimize SVM hyperparameters for fault diagnosis of ball bearings. Section 4 summarizes this paper.

2.Proposed Method

Asymptotic Behaviors of SVM with Gaussian Kernel

Keerthi and Lin took logC and logσ2 as the parameters of two dimensional hyperparameter space and got an interesting property of SVM classifiers with Gaussian kernel: in the hyperparameter space of (logC, logσ2), there exists a contour of generalization error (or an estimate such as leave-one-out error or k-fold cross validation error). It separates the hyperparameter space into two regions: a bad region and a good region. The best hyperparameter combination with the best generalization performance must be in the good region.

Good Region Recognition for SVM Hyperparameters

We put forward a new idea with good region recognition to speed up SVM hyperparameter selection. Distinguishing a region between good or bad can be seen as an easy binary-classification problem, so we can use a SVM classifier to recognize the good region at first. We propose the following method to get a recognition model of good region.

(1) Initialize the two-dimension hyperparameter selection space: [Cmin, Cmax] for C and [σmin, σmax] for σ.

(2) Get a candidate of hyperparameter combinations (C, σ2) in the two-dimension hyperparameter space for the recognition model of good region with a large step r×λ (r is a positive integer called the step scale in this paper, and λ is basic selection step).

(3) Estimate the generalization performance with each candidate of hyperparameter combinations.

(4) According to the Pareto principle , we define that 80th percentile of classification accuracy is the threshold to distinguish good and bad region.

(5) Get a dataset including candidates of hyperparameter combinations and their corresponding classes: good region and bad region.

(6) Train the model for good region recognition using the balanced dataset with a hyperparameter combination (C=1, σ4=2) .

Hyperparameter Selection with Good Region Recognition Method

When searching for the best hyperparameter combination (C, σ2), there is the conventional method called grid search. This is computationally expensive since it needs to train numerous SVM classifiers with all the points in hyperparameter space. In order to overcome this disadvantage, we introduce the proposed good region recognition model for grid search. The process is as follows.

(1) Get the recognition model of good region following the procedure in section 2.

(2) By using the model, classify all the hyperparameter combinations in predefined hyperparameter selection space with the selection step λ to two classes: good region and bad region.

(3) Search for the hyperparameter combination with the best generalization performance within the good region.

When we have a candidate of hyperparameter combinations {(Ci, σi2) | 1≤i≤N}, the proposed method needs to recognize whether each of them is in the good region first. Compared with grid search, it can save time by avoiding model computation of hyperparameter combinations in bad region.

3.Application to Fault Diagnosis

To demonstrate the superiority of the proposed method over grid search, the experiment adopts a dataset: ball bearing vibration data for normal and faulty bearings with rotary speed 1797 rpm. We divide faults data into 12 time records and divide normal data into 24 time records. Totally 27 features are extracted from the experiment data.

We need to test performance of the proposed hyperparameter combination selection method by classifying four kinds of bearing faults. Initially, we use two dimensional hyperparameter selection space: $\exp([-20: r\times\lambda:20])$ for C and $\exp([\log(\sigma1^*)-1: r\times\lambda: \log(\sigma2^*)+2])$ for σ, where $\sigma1^*$ and $\sigma2^*$ are calculated .

We set different selection step scale r values and show results in Table 1. The computation time T consists of two parts: first part of getting recognition model of good region T1, and the second part of getting best hyperparameter combination in good region by grid search T2. Five-fold cross validation is used to estimate SVM classification accuracy. The comparison of computation time between the proposed method and full grid search is show in Figure 1.

Table 1. Experimental Results

r	σ	C	Accuracy	T_1	T_2	T
1.5	6.69	−20	100	289.12	220.89	510.01
2	18.17	−20	100	170.77	223.17	393.94
3	12.18	−2.4	100	70.42	236.52	306.94
4	11.02	−20	100	43.86	233.76	277.62
5	36.60	−20	100	31.09	228.02	259.11
6	12.18	−1.6	100	21.44	234.23	255.67
7	12.18	−3.2	100	16.93	234.17	251.10
8	12.18	−0.8	100	13.15	249.06	262.21
9	4.48	−20	100	9.77	214.10	223.87
10	13.46	−2.4	100	10.04	269.57	279.61
11	11.02	−2.4	100	7.01	293.57	300.58
12	6.05	−20	100	10.22	286.29	296.51
13	5.47	−20	100	4.56	217.93	222.49
14	11.02	−20	100	4.75	221.93	226.68
15	11.02	−1.6	100	4.65	220.93	225.58
16	14.88	−20	100	4.87	252.52	257.39
17	13.46	−2.4	100	2.82	299.85	302.67
18	9.97	−20	100	2.72	298.98	301.70
19	9.97	−20	100	2.87	290.64	293.51
20	14.88	−20	100	2.77	311.86	314.63
21	9.97	−20	100	2.99	304.09	307.08
22	11.02	−2.4	100	2.82	299.76	302.58
23	18.17	−20	100	2.74	296.86	299.60
24	5.47	−20	100	2.83	296.04	298.87
25	13.46	−20	100	3.11	304.15	307.26

According to Figure 1 and Table 1, we can see that the proposed method takes much less time than the grid search for selecting best hyperparameter combination with the same high classification accuracy (100%). For example, when step scale r=9, the proposed method takes the least time (223.87s) that accounts for only 34% in grid search. The boundary of good region and bad region in the recognition model of good region is shown in Figure 2.

Conclusion

Hyperparameter selection is an important research topic to SVM. Inspired by the asymptotic behaviors of Gaussian SVM, we proposes a novel method to find the good region of SVM hyperparameters. The proposed method can speed up optimization search based method. In this paper we have introduced the proposed method to the conventional grid search for finding the best hyperparameters in fault diagnosis of rolling bearings. Experimental results demonstrated its efficiency.

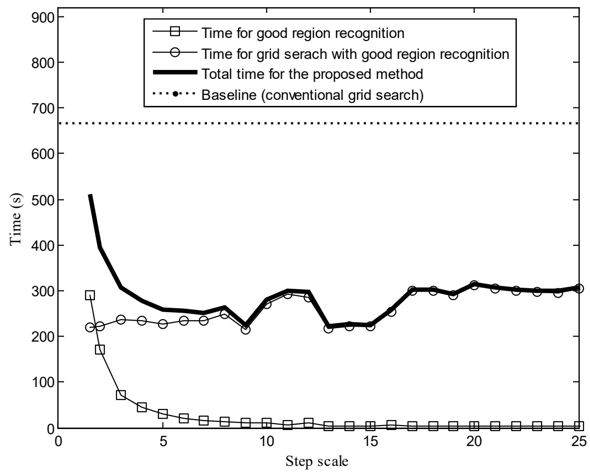

Figure 1. Computation time performance

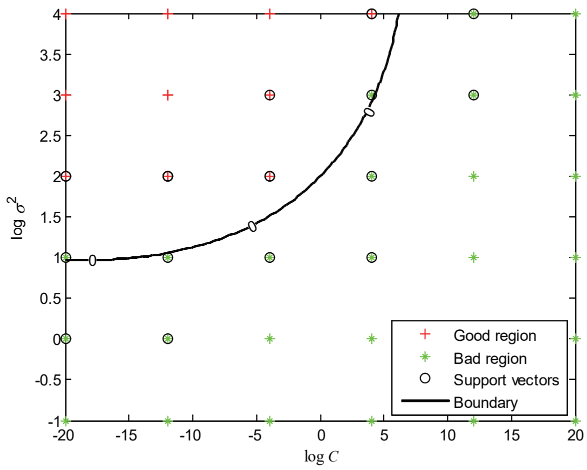

Figure 2. Boundary of good region and bad region when step scale is nine

References

[1]V. Vapnik, The Nature of Statistical Learning Theory. Springer Berlin Heidelberg: Springer Berlin Heidelberg, 2000.

[2]X. H. Liu, P. Cohen, M. Berthold, S. R. Gunn, M. Brown, and K. M. Bossley, "Network performance assessment for Neurofuzzy data modelling," in Advances in Intelligent Data Analysis Reasoning about Data, X. H. Liu, P. Cohen and M. Berthold, Eds. Springer Berlin Heidelberg: Springer Berlin Heidelberg, vol.1280, pp. 313-323, 1997.

[3]D. Lowe and D. Broomhead, "Multivariable functional interpolation and adaptive networks," Complex Systems, vol. 2, pp. 321-355, 1988.

[4]C. Hsu and C. Lin, "A comparison of methods for multiclass support vector machines," Neural Networks, vol. 13, no.2, pp. 415-425, 2002.

[5]S. S. Keerthi and C. J. Lin, "Asymptotic behaviors of support vector machines with Gaussian kernel," Neural Computation, vol. 15, no. 7, pp. 1667-1689, 2003.

[6]Bookstein and Abraham, "Informetric distributions, part I: Unified overview", American Society for Information Science, vol. 41, no.5, pp. 368-375, 1990.

[7]M. Varewyck and J. P. Martens, "A practical approach to model selection for support vector machines with a Gaussian kernel," in Proceeding of IEEE Transactions on Systems, Man, and Cybernetics, pp. 330-340, 2011.

[8]Z. L. Liu, M. J. Zuo, X. Zhao, and H. B. Xu, "An analytical approach to fast parameter selection of Gaussian RBF kernel for support vector machine," Journal of Information Science and Engineering. Accepted on Mar. 11, 2013.