Radar Fault Diagnosis Based on Multiple Kernel Learning SVM

LI Ke-Xin, XIA Hong-Sen, Zhong Ling
Department of Electronic Reconnaissance, Shenyang Artillery Academy, Shenyang, 110162, China
e-mail lekexin@126.com

Abstract: Based on decision tree combined strategy and multiple kernel learning support vector machines，a new fault diagnosis method is proposed to improve precision and speed of radar fault diagnosis. The multi-class classification problem is solved by decision tree combined strategy．And multiple kernel learning support vector machines are able to translate linearly constrained quadratic programming into quadratic ally constrained quadratic programming by mixed kernel space．Simulation results indicate that the diagnosis precision by multiple kernel learning support vector machines is better than the standard support vector machines with less number of support vectors. Moreover，the introduction of decision tree combined strategy guarantees the diagnosis speed improvement．The proposed method can solve the radar fault diagnosis problem accurately and rapidly.

Keywords: fault diagnosis; SVM; multiple kernel learning; decision tree

1 Introduction

It is epoch-making of fault diagnosis to check out system fault rapidly and make self-repaired system rebuild control law in order to avoid system breaking down and equipment loss for these systems of high safety. Fault diagnosis method based on model relies on mostly math's model of system, so becomes unable for fault diagnosis of complicated non-linear system. Nerve network can approach fully free will complicated non-linear relation and be used widely in fault diagnosis area. However nerve network has its defaults, such as exiting local teeny point, beyond learning and relying on experience to structure selection. So its using effect is debased badly.

Differently from general nerve network based on minimum theory of experience risk, support vector machines (SVM) established by Vapnik follows minimum theory of structure risk and integrate several technique ,such as most internal hyper plane, Mercer kernel, protruding quadratic programming, scarcity solution and relaxation variable ,so these can transform non-linear problem into linear problem and obtain optimum solution[1]. The use of SVM in fault diagnosis area is a popularization of multi-class classification arithmetic essentially, so its classification precision ,quantities of support vector and combined strategy of multi-class program effect directly two important factions of fault diagnosis :diagnosis precision and diagnosis rapidly.

This paper adopts decision tree combined strategy that can diagnosis rapidly to realizing faults diagnosis based on SVM, although this method can insure diagnosis rapidly, it can not insure diagnosis precision of decision tree SVM to samples of distributing complicatedly. In recent years, multi-kernel learning develops rapidly, and may be resolve this problem. MKL can transform quadratic programming (QP) problem into semi-definite programming (SDP) problem or quadratic constraint quadratic programming (QCQP) problem of linear restriction of standard SVM, MKL can realize fault diagnosis in mixed kernel space, it can not only advance diagnosis precision ,but also reduce the quantity of support vector, and advance diagnosis speed.

2 Multiple Kernel Learning Support V ector Machines

2.1 Support Vector Machines

Considering a two-value classify question, to the given sample:

$$(x_1, y_1),\ (x_2, y_2),\ \cdots,\ (x_n, y_n) \qquad (1)$$

$$x_i \in R^i,\quad y_i \in \{-1,+1\}$$

Using a non-linear mapping Φ, shining up the data into a multidimensional characteristic space, in this space, we can realize linear classify[2]:

$$f(x_{new}) = \mathrm{sgn}\left(\omega^T \Phi(x_{new}) + b\right) \qquad (2)$$

Where, $\omega \in R^n$, $b \in R$, SVM is upper limit of minishing model generalization error when minimizing sample error, this paper pay more attention to first-order norm soft interval restriction, so we can obtain following optimum problem:

$$\sum_{\omega,\xi}\langle\omega,\omega\rangle + C\sum_{i=1}^{n}\xi_i \qquad (3)$$

Subject to $(\langle w,\Phi(x_i)\rangle + b) \ge 1-\xi_i \quad \xi_i \ge 0$

Where constant C is relaxation variable.

Considering dual problem of (3), we can obtain :

$\omega = \sum_{i=1}^{n}\alpha_i \Phi(x_i)$, α_i is the solution of QP problem [3]:

$$\max_{\alpha} 2\sum_{i=1}^{n}\alpha_i - \sum_{i,j=1}^{n} y_i y_j \alpha_i \alpha_j K(x_i, x_j) \qquad (4)$$

Subject to $0 \le \alpha_i \le C, \sum_{i=1}^{n} y_i\alpha_i = 0 \quad i = 1,\cdots,n$

Where $K(x_i, x_j)$ is the kernel function which is suitable to Mercer condition, (4) can be rewritten as:

$$\max_{\alpha} 2\alpha^T e - \alpha^T diag(y) K diag(y)\alpha \qquad (5)$$

Subject to $0 \le \alpha \le C, \alpha^T y = 0$

Where K is kernel matrix, $e = \{1,1,\cdots,1\}^T$, $y = \{y_1, y_2, \cdots, y_n\}^T$, diag(y) is diagonal matrix.

2.2 Multiple Kernel Learning Support Vector Machines

From reference [4] we know that classify capability of SVM is limited of (5) for kernel matrix trace (trace (K) =c, c is constant) of fixed track [4]. The less of value, the better of SVM classify capability. Corresponding the given kernel matrix in (5), we can think of following mixed kernel, μ_j is weight coefficient.

$$K = \sum_{j=1}^{m}\mu_j K_j \qquad (6)$$

we can obtain optimum problem:

$$\min_{\mu_j}\max_{\alpha} 2\alpha^T e - \alpha^T diag(y)\left(\sum_{j=1}^{m}\mu_j K_j\right) diag(y)\alpha \qquad (7)$$

Subject to $0 \le \alpha \le C, \alpha^T y = 0$

$$\operatorname{trace}\left(\sum_{j=1}^{m}\mu_j K_j\right) = c, \sum_{j=1}^{m}\mu_j K_j \succ= 0$$

Where, $\sum_{j=1}^{m}\mu_j K_j \succ= 0$ shows that $\sum_{j=1}^{m}\mu_j K_j$ is semi-definite.

Through solving the Lagrangian duality problem of (7), we can transform (7) into following SDP problem:

$$\min_{\mu_j, t, \lambda, \nu, \delta} \qquad (8)$$

Subject to $\operatorname{trace}\left(\sum_{j=1}^{m}\mu_j K_j\right) = c, \sum_{j=1}^{m}\mu_j K_j \succ= 0$, $\nu, \delta \ge 0$

$$\begin{bmatrix} diag(y)\left(\sum_{j=1}^{m}\mu_j K_j\right) diag(y) & e+\nu+\lambda y \\ (e+\nu+\lambda y)^T & t - 2C\delta^T e \end{bmatrix} \succ= 0$$

SDP can be essentially thought of as generalized linear programming, which only transforms linear inequality restriction into linear matrix of ordinary linear programming.

Inequality restriction. Although SDP problem of (8) can be solved by inner spot method, complication of arithmetic is far beyond QP problem of (5), so that this method is difficult to extend.

Supposing weight coefficient $\mu_j \ge 0 (j = 1,\cdots,m)$ of (6), from character of kernel function we know $K = \sum_{j=1}^{m}\mu_j K_j$ is symmetry semi-definite matrix, which is effective kernel function. Standardizing kernel matrix $K_j (j = 1,\cdots,m)$ according to (9) [5]:

$$\overline{K}_j\left(x_p, x_q = \frac{K_j(x_p, x_q)}{\sqrt{K_j(x_p, x_p) K_j(x_q, x_q)}}\right) \qquad (9)$$

$$p, q = 1,\cdots,n$$

While transforming SDP problem of (8) into following QCQP problem

$$\max_{\alpha,t} 2\alpha^T e - ct \qquad (10)$$

Subject to $0 \le \alpha \le C, \alpha^T y = 0, t \ge \frac{1}{n}\alpha^T diag(y)\overline{K}_j diag(y)\alpha$

We can solve QCQP problem through optimum software bag MOSEK based on inner spot method, because the arithmetic of QCQP and QP are all complicated and (10) realizes data classification in mixed kernel space obviously, multiple kernel learning SVM is better than standard SVM. When solving (10), we can obtain not only α, t but also weight coefficient μ_j that can show kernel function importance through its duality solution.

Where SVM classify machines of (2) is:

$$f(x_{new}) = \operatorname{sgn}\left(\sum_{j=1}^{m}\sum_{i=1}^{|sv|}\mu_j \alpha_i K_j(x_i, x_{new})\right) \qquad (11)$$

3 Multi-class Classification Support Vector Machines

Fault diagnosis by support vector machines needs designing multiple classification arithmetic, and multi-class classification arithmetic of support vector machines is an extend to dual-class classification problems. Currently multi-class classification arithmetic of support vector machines includes following two: firstly, we can unit inference solution of multi-class classification and obtain an optimization problem; secondly, we can decompose multi-class classification problem into several dual-class classification problem. Because the first method needs using large variables, it is difficult to realize and calculate and its classify precision is not high inequality. So in practice, we generally adopt the second method. According to the differences of combination strategy, the second method can be divided into following: $(1-\alpha, -\mathrm{r})$ and $(1-\alpha, -1)$, DAG and DT.

Support vector machines can be used in fault diagnosis. Its diagnosis speed is not only affected by combination strategy of multiple support vector machines, but also affected by the number of support vector. Obviously, classification strategy of DT is suitable for rapidness of radar fault diagnosis. On one hand, it needs the less number of support vector, on the other hand, its diagnosis strategy can realize diagnosis rapidly. Through multiple kernel learning support vector machines, we can realize linear classification in mixed kernel space, and this can not only advance classification precision, but also minish the number of support vector. So we can realize effectively radar fault diagnosis by using multiple kernel learning support vector machine and classification strategy base on decision tree.

4 Simulation Result and Analysis

Selecting eight radar parameter variables as condition property, including maximum operation range, antenna running range, azimuth searching range, ability to finding target ,tracking distance, angle speed of tracking, distance discrimination, azimuth discrimination. Selecting five sort of these as strategy property, including no fault, filament open circuit of traveling-wave tube, non-stabilization frequency synthesizer, azimuth electromotor marring, elevation electromotor marring. Data of training aggregation is 250,data of each sort is 50;data of testing aggregation is 500,data of each sort is 100. Because actual signal is non-linear, so when simulating, we add white noise whose average value is zero, and variance is $v\bar{f}$ as observation noise to train SVM, $\bar{f}$ is average value of random attribute. Supposing normalization factor C of standard SVM and MKL SVM is 1000, kernel function of standard SVM is Gauss kernel; kernel parameter σ can be obtained from girding searching method. Kernel function of MKL SVM is Gauss kernel too, and the data is two, kernel parameter can be expressed σ_1 , σ_2 ,diagnosis result expressed in chart 1

Table 1 Radar fault diagnosis result

\	MKL SVM				Standard SVM		
	σ_1	σ_2	P/%	$\#SV$	σ	P/%	$\#SV$
0	0.25	0.5	100	40	0.2	98.60	54
0.02	0.25	0.5	98.80	48	0.15	95.20	72
0.05	0.25	0.5	97.40	61	0.15	94.00	95

From Table 1, we know that MKL SVM based on DT diagnosis strategy is obviously better than standard SVM with less number of support vectors, and its speed is faster than standard SVM too. So the proposed method can solve the radar fault diagnosis problem accurately and rapidly.

5 Conclusions

MKL SVM based on DT compositions strategy can solve the radar fault diagnosis problem accurately and rapidly. We can improve diagnosis speed if considering sparseness SVM; this will be studied for the future.

Acknowledgement

Many people have made invaluable contributions, both directly and indirectly to our research. We would like to express our warmest gratitude to Prof HOU Rui-Dong, , for his instructive suggestions and valuable comments on the writing of this thesis. Without his invaluable help and generous encouragement, the present thesis would not have been accomplished. At the same time, we are also grateful to Prof. DIAO Hua-Zong and Prof. ZHOU Hong-Jie for providing us with valuable advice and access to the related resources on my thesis.

References

[1] Vapnik V. The nature of statistical learning theory[M]. New York：Springer—Verlag，1995·

[2] Hsu C．Lin C J．A comparison of methods for mufti—class support vector machines[J]．IEEE Transactions on Neural Networks，2002，13(2)：415-425.

[3] Lanckriet G R G．Bie T D，Cristianini N，et a1．A statistical framework for genomic data fusion[J]．Bioinformatics，2004，20(16)：2626-2635.

[4] Lanckriet G R G．Learning the kernel matrix with semi-definite programming[J]．Journal of Machine Learning Research．2004，5：27=72.

[5] Sonnenburg S．Ratsch G，Sch61kopf B．Large scale nmhiple kernel learning[J]． Journal of Machine Learning Research，2006，7：1531=1565.

978-1-935068-01-3 © 2009 SciRes.