Tech Science Press

Extrapolation for Aeroengine Gas Path Faults with SVM Bases on Genetic Algorithm

Yixiong Yu*

School of Aeronautic Science and Engineering, Beihang University, Beijing, 100083, China
*Corresponding Author: Yixiong Yu. Email: 16051077@buaa.edu.cn.

Abstract: Mining aeroengine operational data and developing fault diagnosis models for aeroengines are to avoid running aeroengines under undesired conditions. Because of the complexity of working environment and faults of aeroengines, it is unavoidable that the monitored parameters vary widely and possess larger noise levels. This paper reports the extrapolation of a diagnosis model for 20 gas path faults of a double-spool turbofan civil aeroengine. By applying support vector machine (SVM) algorithm together with genetic algorithm (GA), the fault diagnosis model is obtained from the training set that was based on the deviations of the monitored parameters superimposed with the noise level of 10%. The SVM model (C = 24.7034; γ = 179.835) was extrapolated for the samples whose noise levels were larger than 10%. The accuracies of extrapolation for samples with the noise levels of 20% and 30% are 97% and 94%, respectively. Compared with the models reported on the same faults, the extrapolation results of the GASVM model are accurate.

Keywords: Aeroengine; extrapolation; gas path fault diagnosis; genetic algorithm; support vector machine

Aeroengine technology is known as the pearl of industry crown and the heart of aircraft. An aeroengine generally operates in a harsh environment, such as at high temperature, high pressure, high vibration, and high noise, which result in its system high fault rate. Further, the aeroengine control system is extremely complex and its maintenance is difficult. The performance of an aeroengine directly affects the security and reliability of the aircraft. There are many kinds of fault modes of aeroengines, among which gas path faults account for a large proportion (≈ 90%) of the total faults. Therefore, it is of great significance to carry out research on fault diagnosis of aeroengines [1, 2].

The clustering analysis based on unsupervised learning theory can be used in aeroengine fault diagnosis. This method evaluates the similarity of parameters, and then classifies the samples [3]. Using supervised learning method to establish fault diagnosis models of aeroengines has become the current research hotspot. Peng has studied the diagnosis principles of 20 faults for double-spool turbofan civil aeroengine with the deviations of measurement parameters (see Tab. 1), and established a diagnosis model for gas path faults [4]. The diagnosis system is based on a radial basis function (RBF) neural network, which consists of five 8-2 sub-networks. The input variables are the deviations of monitored parameters, and the output variables are performance parameters of the aeroengine. The RBF model has good generalization performance. Wang et al. have introduced a fault diagnosis method based on support vector machine (SVM) and synergetic neural network (SNN), in order to distinguish aeroengine gas path faults and improve the diagnosis accuracy [5]. The sample set of parameters measured was expanded at noise level of 10% for 20 aeroengine gas path faults to obtain 200 samples, of which 100 samples were taken as the training set and the remaining 100 as the test set. The SVM model has the diagnosis accuracy of 96%, and the SNN model can distinguish the four similar faults which are misdiagnosed by the SVM model. Du has used hybrid particle swarm optimization (HPSO) combined with twin support vector machine (TWSVM) to diagnose the 20 faults mentioned above [6]. The noise level used is 10% and the number of samples is 200. The prediction accuracies of the training set and the test set are 98% and 97%, respectively. Recently, Jiang et al. have introduced a new method for the diagnosis of aeroengine gas path faults with nine latent variables obtained from 14 monitored parameters, via a combination of principal component analysis (PCA) and deep belief network (DBN) [7]. Based on the PCA and DBN, a fault diagnosis model was established with diagnosis accuracy of 100%, which was higher than that (80%) of DBN model without PCA-based dimensionality reduction.

Apart from the above models dealing with aeroengine gas path faults, Zhang et al. have used the immune algorithm to optimize SVM parameters, and obtained a diagnosis model for aeroengine wear faults [8]. When the noise levels reach to 1%, 2%, 3% and 4%, the diagnostic accuracies are 98%, 96%, 92% and 86%, respectively, which proves that the algorithm is effective. Pi et al. have adopted the genetic algorithm (GA) combined with SVM to develop a diagnosis model for aeroengine wear faults, which has an accuracy of 98.21% [9]. Cao et al. have used a combination of fuzzy rough set and SVM to establish a diagnosis model of aeroengine faults, with the accuracy being 95% [10]. Cao et al. have applied AdaBoost algorithm to build a fault diagnosis model of aeroengines [11]. The diagnostic accuracy is 97.3%, which illustrates that the AdaBoost algorithm can significantly improve the performance of the classifier. Cui et al. have developed a fault diagnosis model for a main fuel pump of aeroengines by applying the wavelet packet transform and an extreme learning machine (ELM) [12]. The diagnostic accuracy is 96.67%.

All the models mentioned above are based on the interpolation method to predict the samples in test sets, that is, the range of variables from the test sets is close to that of training sets. Because of the complexity of working environment and faults of aeroengines, it is unavoidable that the monitored parameters vary greatly, which mean that a fault diagnosis model for aeroengines should have good extrapolation diagnosis ability. Extrapolation is process of estimating the value of a variable or function outside the tabulated or observed range. The purpose of this study is to analyze and model 20 gas path faults of a double-spool turbofan civil aeroengine. The samples in the training set are based on the deviations of the monitored parameters with the noise level of 10%. From the training set, a diagnosis model of gas path faults was developed with GASVM algorithm, which was extrapolated for the samples whose noise levels were larger than 10%.

1 Modeling Method

The SVM algorithm has many attractive characteristics and advantages. Unlike the traditional theory, SVM adopts the principle of structural risk minimization (SRM), which makes it possible for SVM to have good prediction ability even if training samples are shortage [13-16]. SVM was originally used in pattern recognition. With the introduction of insensitive loss function, the application of SVM extends to the prediction in nonlinear regression and to the evaluation in time-series analysis. For n-dimensional training samples (x_i; y_i), $i = 1, \ldots, l$; $x_i \in R^n$; $y \in \{1, -1\}$, 1 and -1 represent the dependent variables. The optimal hyperplane problem to be solved by SVM is expressed as:

$$\min_{w,\xi,b} J(w, \xi, b) = \frac{1}{2} w^T w + C \sum_{l}^{i} \xi_i \tag{1}$$

$$y_i(\mathrm{w}^T\varphi(x_i)+b)\geq 1-\xi_i \tag{2}$$

$$\xi_i\geq 0 \tag{3}$$

Here C is a penalty parameter used for adjusting the training error. w and b are coefficients estimated. To construct an optimal linear classification surface, the training data set $\mathbf{x}_i$ is mapped to the high-dimensional feature space by the kernel function. Thus, the optimization problem becomes as follows:

$$maxW(\alpha)=\sum_i\alpha-\frac{1}{2}\sum_{ij}\alpha_i\alpha_jy_iy_jK\left(x_i,x_j\right) \tag{4}$$

$$\sum_i\alpha_\mathrm{i}y_i=0 \tag{5}$$

$$0\leq\alpha_i\leq C \tag{6}$$

Where α_i and α_i^* are the Lagrange multipliers that are used to solve the quadratic optimization problem; and $K(x_i,x_j)$ is the kernel function realizing nonlinear transform. The following optimal classification function can be obtained in solving the equation above.

$$f(x)=sgn\left[\sum_{i=1}^n\alpha_iy_iK(x,\ x_i)+b\right] \tag{7}$$

In present work, the following Gaussian radial basis function (RBF) was used:

$$K(\mathrm{x}_i,\mathrm{x}_j)=\exp(-\gamma\left\|\mathrm{x}_i-\mathrm{x}_j\right\|^2) \tag{8}$$

Here the parameter γ denotes the kernel width. The SVM parameters C and γ that affect the fitting degree and generalization ability of SVM are tuned with the GA method. GA simulates the biological evolution and genetic mechanism of the biological world and forms a search algorithm for global optimization, which provides an effective method for SVM parameter selection. GA first generates the initial population. The fitness function is then used for the operations on individual in the heredity, including selection, crossover and mutation, to retain the individual with good fitness value, and to eliminate that with poor fitness value. Thus, the new population is produced, and evolves again, until the most optimized solution or the most reasonable solution is obtained. The flow chart of GASVM algorithm is shown in Fig. 1.

2 Case Analysis of Gas Path Fault Diagnosis for Aeroengines

The data samples in this paper are from reference [4] (see Tab. 1). The monitored parameters are as follows: flight altitude H = 10700 m; atmospheric pressure p_0 = 0.23723 bar; atmospheric temperature T_0 = 218.6 K; atmospheric density ρ_0 = 0.37806 kg/m; flight Mach number Ma = 0.395518; and the thrust F_N = 47.01 kN. The monitored parameters include high-pressure rotor speed, δN_1; low-pressure rotor speed, δN_2; fan pressure ratio, $\delta\pi_\mathrm{F}$; boost stage pressure ratio, $\delta\pi_\mathrm{LC}$; compressor pressure ratio, $\delta\pi_\mathrm{HC}$; total inlet temperature of high-pressure compressor, δT^*_{25}; exhaust temperature of low-pressure turbine, δT^*_5; and fuel consumption, δwf.

In this paper, the deviations of measurement parameters under various fault modes in Tab. 1 are superimposed with noise signals according to Eq. (9), so that each fault is superimposed to produce 10 samples, and 20 faults form 200 samples.

$$x_\mathrm{i}=x_\mathrm{i}\ (1+L\times\sigma_\mathrm{fault}\times randn) \tag{9}$$

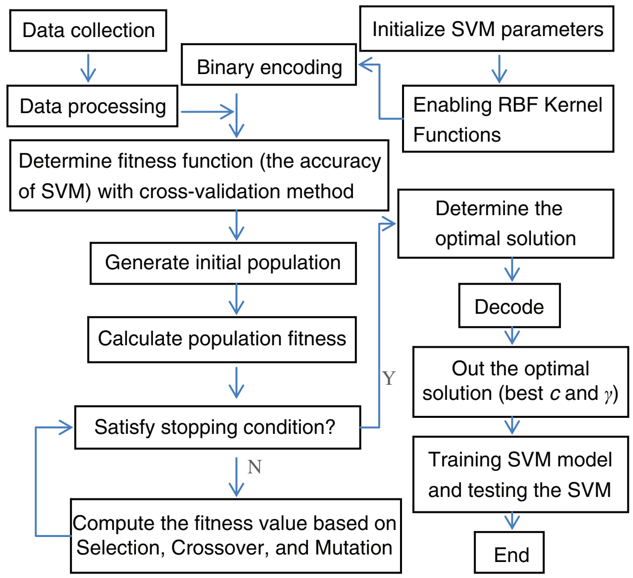

Figure 1: Flow chart of GASVM algorithm

Table 1: Diagnosis principles of 20 gas path faults for a double-spool turbofan civil aeroengine

No.	Symbol	δN_1 (%)	δN_2 (%)	$\delta \pi_F$ (%)	$\delta \pi_{LC}$ (%)	$\delta \pi_{HC}$ (%)	δT^*_{25} (%)	δT^*_5 (%)	δWf (%)
1	F1	3.59	3.06	–1.27	0.27	4.78	1.93	4.03	7.16
2	F2	1.58	1.74	–0.66	0.56	1.82	0.89	1.33	2.65
3	F3	3.37	–0.62	–0.51	–3.56	5.96	0.32	3.15	4.49
4	F4	1.86	–0.29	–0.28	–1.73	2.94	0.21	1.22	1.81
5	LC1	3.58	–0.02	0.56	–8.97	7.14	–0.44	3.45	1.68
6	LC2	2.39	0.00	0.30	–5.35	4.60	–0.33	1.61	0.88
7	LC3	0.71	–0.10	–0.20	0.20	0.45	0.96	1.10	1.05
8	LC4	0.35	–0.03	–0.12	0.12	0.32	0.38	0.40	0.47
9	HC1	1.11	0.00	0.26	3.66	–4.22	0.30	1.63	1.12
10	HC2	1.85	0.01	0.04	0.65	–0.66	–0.03	0.17	0.02
11	HC3	–5.86	–0.06	0.25	7.94	–9.30	0.95	3.60	2.08
12	HC4	–2.30	0.06	0.22	2.62	–3.18	0.19	1.12	0.84
13	HT1	1.85	–0.03	–0.07	–0.85	7.59	0.03	–0.15	0.08
14	HT2	–2.77	–0.03	0.10	2.49	–8.42	0.11	1.05	0.57
15	HT3	–5.42	–0.11	0.18	6.17	–12.77	0.56	2.72	1.50
16	HT4	–6.95	–0.20	0.33	9.41	–11.83	1.21	5.12	2.96
17	LT1	–4.65	–0.06	0.24	5.66	–7.24	0.52	2.27	1.27
18	LT2	4.88	–0.38	–0.56	–5.05	8.56	–0.52	1.71	2.22
19	LT3	5.71	–0.72	–0.97	–6.27	10.86	–0.87	4.48	4.66
20	LT4	3.54	–0.90	–1.11	–4.02	7.49	–0.63	6.58	6.19

Here *randn* is a random function obeying normal distribution and σ_{fault} is the standard deviation of parameters measured in Tab. 1. In this paper, noise signal L is equal to 10%, 20%, 30%, 40%, 50%, or 60%.

After adding 10% noise signal to the deviation of each measurement parameter in Tab. 1, 200 samples were obtained, which were divided into two sets. One was used as the training set to build a model, and the other was used as the test set to evaluate the model. The LIBSVM toolbox is used on MATLAB R2014a platform for developing SVM models. In the process of GASVM modeling, the parameters are set as follows: the maximum evolutionary algebra is 200; the maximum population number is 20; the leave-one-method (LOO) is adopted for the cross validation; the search range of SVM parameter C is 20-100; and the search range of parameter γ is 100-200. In the end, the optimum SVM parameters, C = 24.7034 and γ = 179.835 were obtained.

The prediction accuracy of the training set based on the optimal SVM is 100%, and the model was evaluated by the test set. The prediction accuracy from the test set is 99%. The prediction results are shown in Fig. 2. The sample set (see Tab. 1) has been studied by other research groups [5, 6], with the same numbers of samples for the training set and test set. The prediction accuracy of SVM model from reference [5] is 96%; and the prediction accuracies of training set and test set in reference [6] are 98% and 97%, respectively. Therefore, compared with references [5, 6], the GA-SVM model in this paper is accurate. However, it should be noted that the parameters of SVM (C = 24.7034; γ = 179.835) in this model are obviously larger than those in references [5, 6] (C < 16; γ < 28). Generally, when the SVM parameters are too large, the SVM models tend to produce over-fitting, that is, the prediction accuracy of training set is high, while that of test set is low. In fact, the prediction accuracy from the test set is 99%, which is almost consistent with the prediction results of the training set. Therefore, there is no risk of over-fitting for SVM models.

Figure 2 shows that only one case of mechanical damage of high-pressure turbine blades (HT4) is misdiagnosed as low-pressure turbine blade scaling (LT1) by GASVM model in this paper. The prediction results show that the fault types HT4 and LT1 are similar, which are consistent with the conclusions of Wang [5].

Because of the complexity of working environment and faults of aeroengines, it is unavoidable that large deviations of monitored parameters occur. Therefore, the GASVM model in this paper was extrapolated for five noise levels of 20%, 30%, 40%, 50% and 60%. The diagnosis accuracies are 97.0%, 94.0%, 89.0%,

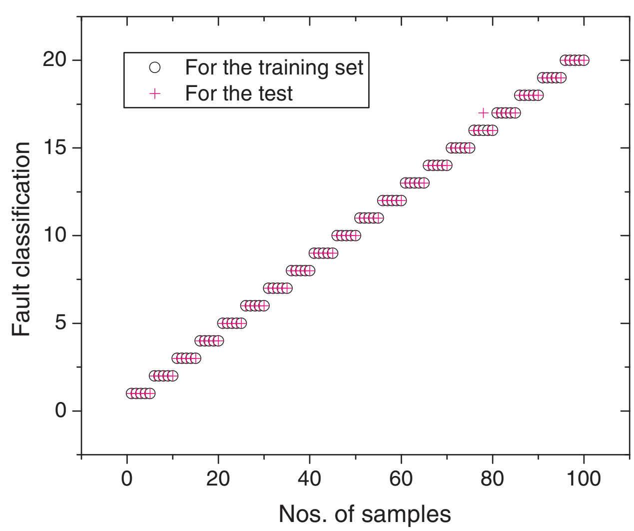

Figure 2: Prediction results of fault classification based on the noise level of 10%

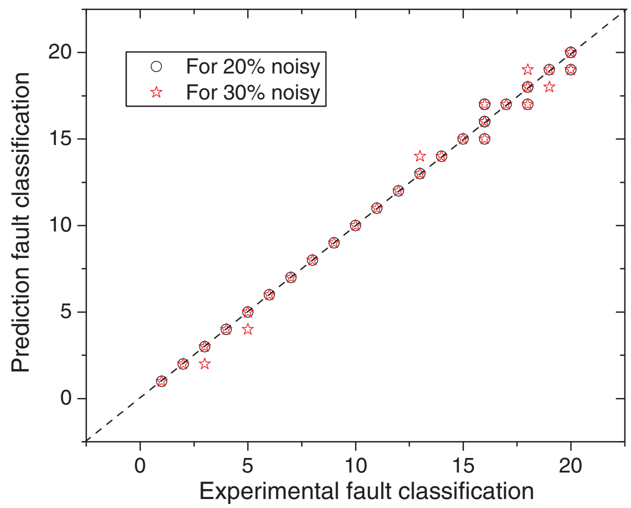

Figure 3: Prediction results of fault classification for samples with high noise levels

88.0% and 84.5% respectively. The extrapolation diagnosis results from 20% and 30% noise levels are shown in Fig. 3. Compared with references [5, 6], the GASVM model is successfully extrapolated for the samples with 20% noise level, which is equal to the prediction accuracy of references [6] (97.0%). The prediction accuracy (94.0%) of extrapolation for 30% noise level is also close to that of reference [6]. Therefore, the GASVM model obtained from the training set with the noise level of 10% can be extrapolated for the samples in the test with the noise levels of 20% and 30%. It should be pointed out that the extrapolated range of our GASVM is only applicable for samples expanded from Eq. (9) although there are various ways of noise signal superposition.

3 Conclusions

In this paper, the GASVM algorithm is used to develop a fault diagnosis model for 20 gas path faults of a double-spool turbofan civil aeroengine by applying the training set of samples with deviations of monitored parameter superimposed a noise level of 10%. The optimal parameters of SVM model are $C = 24.7034$; $\gamma = 179.835$. The prediction accuracies of the training set and test set are 100% and 99% respectively. Moreover, the model was extrapolated for the samples with the noise levels of 20%, 30%, 40%, 50% and 60%. The diagnosis accuracies are 97.0%, 94.0%, 89.0%, 88.0%, 84.5% respectively. The extrapolated prediction results for the noise levels of 20% and 30% are equal to or close to that of previous models. Therefore, the GASVM model in this paper can be used for fault diagnosis of the double-spool turbofan civil aeroengine with the noise level less than 30%.

Acknowledgement: This work was supported by the National Natural Science Foundation of China (41701440).

References

1. Ai, J. L., Yang, X. Z. (2017). Fault diagnosis of aeroengine based on self-adaptive neural network. *Scientia Sinica Technologica* (in Chinese), *48(3),* 326–335. DOI 10.1360/N092017-00224.
2. Palacios, A., Martínez, A., Sánchez, L., Couso, I. (2015). Sequential pattern mining applied to aeroengine condition monitoring with uncertain health data. *Engineering Applications of Artificial Intelligence, 44,* 10–24. DOI 10.1016/j.engappai.2015.05.003.

3. Wang, Y., Zhang, Z., Qian, W., Xu, C., Nie, G. (2018). Research on aero engine fault prediction based on fuzzy clustering. *Journal of Nanchang Hangkong University: Natural Sciences* (in Chinese), *32(1),* 23–28. DOI 10.3969/j.issn.1001-4926.2018.01.004.
4. Peng, S. (2012). *Study on the Gas Path Fault Diagonseis Technology of Aeroenginese*. Shanghai: Shanghai Jiao Tong University (in Chinese).
5. Wang, X., Li, C., Gao, M., Li, Z. (2014). Fault diagnosis of aeroengine gas path based on SVM and SNN. *Journal of Aerospace Power* (in Chinese), *29(10),* 2493–2498. DOI 10.13224/j.cnki.jasp.2014.10.029.
6. Du, Y. (2016). *Aircraft Engine Fault Diagnosis Based on Twin Support Vector Machines*. Nanjing: Nanjing University of Aeronautics and Astronautics (in Chinese).
7. Jiang, L., Li, W., Cui, J., Yu, M., Wang, J. (2019). Fault diagnosis of aeroengine gas path system based on PCA and DBN. *Journal of Shenyang Aerospace University* (in Chinese), *36(1),* 57–62. DOI 10.3969/j.issn.2095-1248.2019.01.009.
8. Zhang, J., Li, Y., Cao, Y., Zhang, L. (2017). Immune SVM used in wear fault diagnosis of aircraft engine. *Journal of Beijing University of Aeronautics and Astronautics* (in Chinese), *43(7),* 1419–1425. DOI 10.13700/j.bh.1001-5965.2016.0553.
9. Pi, J., Ma, S., He, J., Kong, Q. Ma, L. 2018. Application of genetic algorithm optimized SVM in aeroengine wear fault diagnosis. *Lubrication Engineering* (in Chinese), *43(10),* 89–97. DOI 10.3969/j.issn.0254-0150.2018.10.016.
10. Cao, Y., Zhang, J., Li, Y., Zhang, L. (2017). Aeroengine fault diagnosis based on fuzzy rough set and SVM. *Journal of Vibration, Measurement & Diagnosis* (in Chinese), *37(1),* 169–173. DOI 10.16450/j.cnki.issn.1004-6801.2017.01.027.
11. Cao, H., Gao, S., Xue, P. (2018). Aeroengine fault diagnosis based on multi-classification AdaBoost. *Journal of Beijing University of Aeronautics and Astronautics* (in Chinese), *44(9),* 1818–1825. DOI 10.13700/j.bh.1001-5965.2017.0774.
12. Cui, J., Liu, H., Tao, S., Yu, M., Gao, Y. (2018). Aeroengine fault diagnosis method based on ELM. *Fire Control & Command Control* (in Chinese), *43(4),* 113–116. DOI 10.3969/j.issn.1002-0640.2018.04.025.
13. Seo, D.-H., Roh, T.-S., Choi, D.-W. (2009). Defect diagnosis of gas turbine engine using hybrid SVM-ANN with module system in off-design condition. *Journal of Mechanical Science and Technology, 23(3),* 677–685. DOI 10.1007/s12206-008-1120-3.
14. Fatemi, M. H., Ghareh Chahi, Z. (2012). QSPR-based estimation of the half-lives for polychlorinated biphenyl congeners. *SAR and QSAR in Environmental Research, 23(1–2),* 155–168. DOI 10.1080/1062936X.2011.645876.
15. Yu, X., Yu, Y., Zeng, Q. (2014). Support vector machine classification of streptavidin binding aptamers. *PLOS ONE, 9(6),* e99964. DOI 10.1371/journal.pone.0099964.
16. Chang, C. C., Lin, C. J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology, 2(3),* 27. DOI 10.1145/1961189.1961199.