A novel Bayesian Least Squares Support Vector Machine based Anomaly Detector for Fault Diagnosis

Taimoor Khawaja[1], Dr. George Vachtsevanos[2]

[1,2] *Intelligent Control Systems Laboratory, Georgia Institute of Technology, Atlanta, GA, 30080, USA*

taimoor@gatech.edu
gjv@ece.gatech.edu

ABSTRACT

Anomaly detection is the identification of abnormal system behavior, in which a model of normality is constructed, with deviations from the model identified as “abnormal”. Complex high-integrity systems typically operate normally for the majority of their service lives, and so examples of abnormal data may be rare in comparison to the amount of available normal data. Anomaly detection is particularly suited for Intelligent Fault diagnosis of such systems since it allows previously-unseen or poorly-understood modes of failure to be correctly identified. In this paper, we propose a novel Least Squares Support Vector Machine (LSSVM) based Anomaly Detector for efficiently and accurately detecting imminent faults in complex non-linear systems. The Anomaly Detector is supplemented with a Bayesian Inference Framework in order to allow for a probabilistic interpretation of the classification results. Experiments conducted on data from real test cases discussing crack growth on a planetary gearplate on board a UH-60 BlackHawk Aircraft and bending fan blades aboard a chiller show that the Bayesian LSSVM (B-LSSVM) Anomaly Detector can give high identification rates for both the prescribed ‘unknown’ fault samples and the known fault samples.

1. INTRODUCTION

Condition monitoring and diagnosis of machinery is an important field of engineering study (Antoni, 2002). It includes signal measurement, feature extraction, condition recognition and diagnosis decision-making. In substance, condition monitoring is a pattern recognition or classification problem which typically requires training data including samples both from healthy and faulted systems. It is possible, however, that novel faults are evolving while a monitored machine is running. These faults are different from those that have been trained to the diagnostic system and need to be promptly detected. Therefore, it is desirable that such diagnostic systems should not only correctly discriminate all trained faults, but also detect unseen faults. For a machine learning system the ability to identify a novel pattern class is known as novelty detection (Markou, 2003). If this “novel” data deviates from a well defined notion of “normal behavior”, the process is referred to as Anomaly detection.

Statistical pattern recognition techniques, including density estimation, k-nearest neighbor algorithm and Artificial Neural Networks (ANNs) have been explored extensively for machinery fault classification and diagnosis in the last decade. One of the simplest approaches to anomaly detection is based on thresholding output of Multi-layer perceptrons and radial basis function networks (Stephano, 2000). While ANNs are useful black box approaches, capable of approximating any continuous function without assuming any hypothesis about the underlying model, they suffer from the problems of local optima and difficulty in solution interpretation in traditional analytic terms and thus require extensive training data and training time.

An alternate approach to anomaly detection is to define the “normal” data through bounded regions that contain (almost) all data, and use restricted shapes such as hyperspheres for their class boundaries. Support Vector Machine (SVM) classifiers based on the principles of Structural Risk Minimization (SRM) seem to give a flexible and tight data description among these boundary approaches (Tax, 2004). Unlike ANNs, the solution to an SVM classification is well interpreted, and unique. The computational complexity does not depend on the dimensionality of the input space and the SRM principle makes them less prone to over-fitting. For these reasons, SVMs often outperform ANNs and have been successfully applied to many practical problems in recent years (Dehmeshki, 2004), (Soman, 2003), (Hu, 2005).

A one-class classifier for anomaly detection was first proposed by Scholkopf and Smola (Schölkopf, 2000). With this algorithm, one can compute a set of contours enclosing the data points by estimating an optimal separating hyperplane. These contours can be considered as normal data boundaries. The data outside the boundaries are interpreted as anomalies. Many reformulations of the problem have since been derived in an attempt to improve the performance of Scholkopf's Novelty Detector. Ming-Qing Pan et al., for instance, derived the Support vector data description (SVDD) single class classifier based on the premise that as opposed to a hyper-plane, an optimal (minimal volume) sphere containing all the objects belonging to the base class is sufficient (Pan, 2005).

Algorithmically, SVM formulations lead to a quadratic programming (QP) problem, which has the important advantage that the obtained solution is always globally optimal. However, solving a QP problem also implies using more computational time. Indeed the size of the matrix in the QP problem is directly proportional to the number of training samples which limits the application of SVM classifiers in real-time applications. Furthermore, SVMs produce an uncalibrated value which is thresholded to obtain a binary classifier. As such, Novelty Detectors based on SVMs also produce a 0/1 output similar to traditional classifiers. As notions of uncertainty, vagueness, randomness evolve and methods for quantification and management of these notions mature, the utility of designing a classifier that produces a posterior probability P(class) at its output becomes evident specially in the case of modern diagnostics and prognostics systems. For example, a posterior probability allows decisions that are based on a user-specified confidence level. Posterior probabilities are also required when a classifier is making a small part of an overall decision and the classification outputs must be combined for the overall decision.

In this paper, we propose a comprehensive scheme for detecting Anomalous behavior in complex non-linear systems based on a Least Squares formulation for SVMs (LSSVM) proposed by Suykens et. al. (Suykens, 1999). The scheme reformulates the SVM problem such that the solution requires solving a Linear Programming (LP) problem as opposed to the QP problem thereby relaxing the computational burden for huge datasets. Additionally, a Bayesian Inference scheme designed for the LSSVM (Van Gestel, 2002) classifier is adapted and incorporated in the framework so that the Anomaly Detector outputs a posterior probability for the detected class. The rest of the article is organized as follows. We start with a brief discussion of LSSVMs and Bayesian Inference modeling in Section II. The proposed LSSVM Novelty Detector and its Bayesian extensions are explained in Section III. In Section IV, vibration data acquired from a growing crack on a planetary gearplate and proximity data from the fan blades aboard a chiller are analyzed using this scheme and the diagnosis results are presented. Finally, concluding remarks and future research vision is presented in Section V.

2. LEAST SQUARES SVM AND BAYESIAN INFERENCE

Support vector machines are a type of hyperplane classifier which attempt to find an optimal separating hyperplane for binary classification of the input training data given by the following Equation:

$$y(x) = sign[w^T \varphi(x) + b] \tag{2-1}$$

where w is a d-dimensional vector and b is a scalar. The non-linear function $\varphi(.)$ maps the input space to a so-called higher dimensional feature space where the classification is assumed to be linearly separable. When formulated under SVMs the optimal hyperplane is the decision boundary that attains the maximum margin of separation between the two (linearly separable) classes in the feature space. According to the structural risk minimization principle, finding a hyperplane such that the risk bound is minimized is cast as the following quadratic optimization problem (Boser, 1992):

$$\min_{(w,\xi_i)} J(w,\xi_i) = \min \left\{ \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l} \xi_i \right\}$$
$$s.t. \quad \begin{cases} y_i\left(\left\langle w, \varphi(x_i)\right\rangle + b\right) \geq 1 - \xi_i \\ \xi_i \geq 0 \end{cases} \tag{2-2}$$

The positive real constant C should be considered as a tuning parameter in the algorithm. The variables ξ_i are slack variables which are needed in order to allow misclassifications in the set of inequalities (e.g., due to overlapping distributions). Finally, *$\varphi(x)$* represents a non-linear function that maps the input space into a higher dimensional kernel space where a linear separation is exhibited and separation margins, *M*, can be drawn. This is illustrated in *Figure 2-1*. The optimization problem given in *Equation (2-2)* is solved in its dual Lagrangian form the resulting classifier is evaluated in its dual representation.

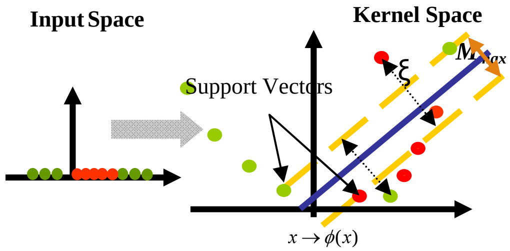

Figure 2-1 A linearly non-separable two-class problem with optimal margin, M_{max}

LSSVMs are an efficient reformulation of the traditional SVMs which solve a Linear Programming (LP) problem as opposed to a Quadratic Programming (QP) problem in the dual Lagrangian form. The LSSVM defines a least squares cost function and replaces the inequality constraints in *Equation (2-2)* with equality constraints (Suykens, 1999) as shown in *Equation (2-3)* for the case of a binary classifier:

$$\begin{aligned} &\min_{w,b,e_k} J = \frac{\mu}{2}\|w\|^2 + \frac{\zeta}{2}\sum_{k=1}^{l} e_k^2 = \mu E_W + \zeta E_D \\ &s.t. \quad y_k\left[\left\langle w, \varphi(x_k)\right\rangle + b\right] = 1 - e_k \\ &\qquad \text{where} \\ &e_k = y_k - \left(\left\langle w, \varphi(x_k)\right\rangle + b\right) \end{aligned} \tag{2-3}$$

Both μ and ζ should be considered as hyperparameters in order to tune the amount of regularization versus the sum squared error (e_k). The primal problem stated above is transformed into its Lagrangian dual and can be simplified to an elegant solution based on a set of linear equations given in *Equation (2-4)*.

$$\begin{bmatrix} 0 & -\overline{Y}^T \\ \overline{Y} & \overline{ZZ^T + \gamma^{-1} I} \end{bmatrix} \begin{bmatrix} b \\ \overline{\alpha} \end{bmatrix} = \begin{bmatrix} 0 \\ \overline{1} \end{bmatrix} \tag{2-4}$$

where $\gamma = \zeta / \mu$, $Y = [y_1;\ldots;y_N]$, $\overline{1} = [1;\ldots;1]$, $e = [e_1;\ldots;e_N]$ and $Z = [\varphi(x_1)^T y_1;\ldots;\varphi(x_N)^T y_N]$. where $\alpha_i \in R$ are the Lagrange multipliers. In short, LSSVMs are reformulations to the standard SVMs which lead to solving simpler linear programming problem. LSSVMs are closely related to regularization networks and Gaussian Processes but additionally emphasize and exploit primal-dual interpretations. Links between kernel versions of classical pattern recognition algorithms such as kernel Fisher discriminant analysis and extensions to unsupervised learning, recurrent networks and control are available (Van Gestel, 2002).

Bayesian learning methods have been successful for the training and understanding of classical neural networks (Bishop, 1995), (MacKay, 1992). In (Van Gestel, 2000), a Bayesian framework has been developed for LSSVMs with three levels of inference. At the first level of inference one considers a probability distribution on w (in a potentially infinite dimensional space) where the prior corresponds to the regularization term $w^T w$ and the least squares cost function to the likelihood. This interpretation allows for probabilistic interpretations of the LSSVM. At the second level of inference one infers the hyperparameters $\gamma = \frac{\zeta}{\mu}$. Finally, at the third level of inference one obtains model comparison criteria after computation of the Occam factor. The kernel width σ is selected at this level.

In the following section, we derive a one-class classifier based on the LSSVM which can be easily incorporated as a diagnostician in an online PHM system. The application of Bayes rule allows us to evaluate the class posterior probability $p(y \mid x)$ and also allows the algorithm to self tune itself by adjusting the hyper-parameters μ and ζ .

3. BAYESIAN LSSVM ANOMALY DETECTOR

The proposed LSSVM Anomaly Detector modifies the LSSVM formulation given in *Equation (2-3)* in order to make it feasible for one-class classification problem. The modified LSSVM problem is thus given by *Equation (3-1)*:

$$\begin{aligned} &\min_{w,\rho,e_k} J_d = \frac{1}{2}\|w\| + \frac{\gamma}{2}\sum_{k=1}^{N} e_k^2 - \rho \\ &s.t. \ \left\langle w, \varphi(x_k)\right\rangle = \rho - e_k \end{aligned} \tag{3-1}$$

Similar to the original LSSVM formulation, we have equality constraints which result in a set of linear equations similar to *Equation (2-4)*. The motivation for such a formulation is derived from Chen and Scholkopf's paper on Novelty Detection using SVMs (Chen, 2005) where the goal is to try to estimate a function $f(x)$ which is positive on the base class and negative on the complement. The form of $f(x)$ is given by a kernel expansion in terms of a potentially small subset of the training data; it is regularized by controlling the length of the weight vector in an associated feature space. This function takes the value +1 in a "small" region capturing most of the training data points, and -1 elsewhere. This is accomplished by mapping the data into the feature space corresponding to the kernel, and to separate them from the origin with maximum margin.

For a new point x, the value $f(x)$ is determined by evaluating which side of the hyperplane it falls on, in feature space. Via the freedom to utilize different types of kernel functions, this simple geometric picture corresponds to a variety of nonlinear estimators in input space. Formulating the classification scheme in the dual space and minimizing the cost function according to Appendix A, the solution can be written in matrix form as a set of linear equations as follows:

$$\begin{bmatrix} 0 & \vec{1} \\ -\vec{1} & \Psi + \gamma^{-1} I \end{bmatrix} \begin{bmatrix} \rho \\ \vec{\alpha} \end{bmatrix} = \begin{bmatrix} 1 \\ \vec{0} \end{bmatrix} \qquad (3\text{-}2)$$

where $\Psi_{kl} = \varphi(x_k)^T \varphi(x_l)$. The Lagrange multipliers α_k are proportional to the errors at the data points so that the solution is no longer sparse as in the original SVM formulation.

A probabilistic framework has been related to the LSSVM classifier, which, under the assumption of a separable Gaussian prior independent of ζ and a Gaussian distribution for the errors e_k, suggests the following expression for the posterior distribution:

$$\begin{aligned} p(w,b \mid D, \log\mu, \log\zeta, H) &\propto \exp\left(-\frac{\mu}{2} w^T w - \frac{\zeta}{2} \sum_k e_k^2\right) \\ &= \exp\left(-\left(\mu E_w + \zeta E_D\right)\right) \end{aligned} \qquad (3\text{-}3)$$

The maximum a posteriori estimates w_{MP} and b_{MP} are then obtained by minimizing the negative logarithm of *Eqation (3-3)*. In the dual space, this corresponds to solving the linear set of equations given by *(3-2)*. Given the posterior probability of the model parameters w and b, we can then integrate over all w and b values in order to obtain the posterior class probability $P(y \mid x, D)$ according to Bayes Rule as follows:

$$p(y \mid x, D) \propto p(y) p(x \mid y, D)$$

where

$$p(x \mid y, D) = \frac{1}{\sqrt{2\pi\left(\zeta_+^{-1} + \sigma_{e_+}^2\right)}} \exp\left[-\frac{m_{e_+}^2}{2\left(\zeta_+^{-1} + \sigma_{e_+}^2\right)}\right] \qquad (3\text{-}4)$$

is the likelihood and $p(y)$ is the prior

where

$$m_{e_+} = w_{MP}^T\left(\varphi(x) - \hat{m}_+\right) = \sum_k \alpha_k K(x, x_k) - \hat{m}_{d_+}$$

and

$$\hat{m}_{d_+} = 1/N_+ \sum_k \alpha_k \sum_{i \in +} K(x_k, x_i) \qquad (3\text{-}5)$$

The term ζ_+^{-1} represents the variance from the target noise while the additional variance $\sigma_{e_+}^2$ is due to the uncertainty in the parameters w. The reader is encouraged to refer to (Van Gestel, 2002) for detailed derivations of the expressions. Results presented in the next section suggest that the Bayesian LSSVM Anomaly Detector is simple, efficient and has excellent generalization capabilities.

4. TEST CASE: APPLICATION TO PLANETARY GEARPLATE VIBRATION DATA

A UH-60 Blackhawk gearbox with a growing axial crack fault on the gearplate was chosen as a real-world test case at the Intelligent Controls System Laboratory (Wu, 2004). The research team designed a Finite Element ANSYS model of the plate, generated artificial vibration data based on the model and inferred several features from it which could reflect the growth pattern of a simulated fault. An overview of the system Finite Element Model (ANSYS) and its equivalent mechanical layout is shown in *Figure 4-1*.

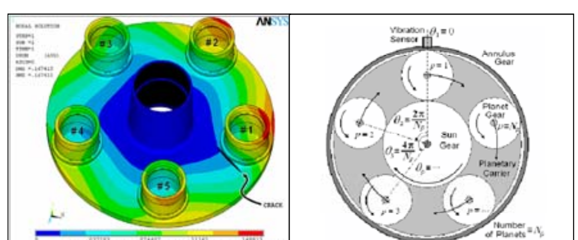

Figure 4-1 (a) ANSYS model of the gearbox plate (b) Mechanical layout

A developing crack close to one of the planetary gears as shown in *Figure* 4-1(a) can lead to a critical failure condition in the aircraft. With the purpose of testing the feasibility and efficiency of such algorithms a seeded fault test was conducted to collect fault data under a fixed known loading profile. A ground-air-ground cycle (GAG) is defined to accommodate different operating conditions. For testing purposes, GAG cycles correspond to time samples.

Raw vibration signals were first denoised to get rid of artifacts and environmental noise (Zhang and Khawaja, 2007). Features identified during the modeling phase were used to extract fault growth patterns from the denoised data. The Sideband Ratio (SBR) feature is well correlated with the actual fault evolution and is therefore selected in this research work. The feature varying over time (GAG Cycles) and the training samples are highlighted in *Figure 4-2*.

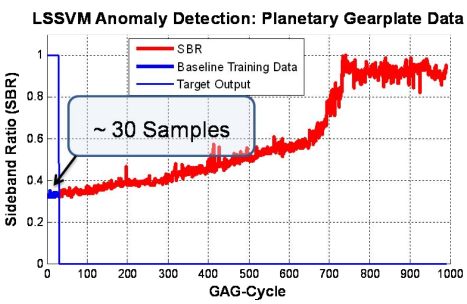

Figure 4-2 Feature correlated with evolving fault. 30 samples of initial feature data are used for training

Thresholded and Probabilistic results from the detection module are given in Figure 4-4 and Figure 4-4. With the proposed LSSVM Anomaly Detector, a fault is detected with 100% accuracy around GAG-cycle 124.

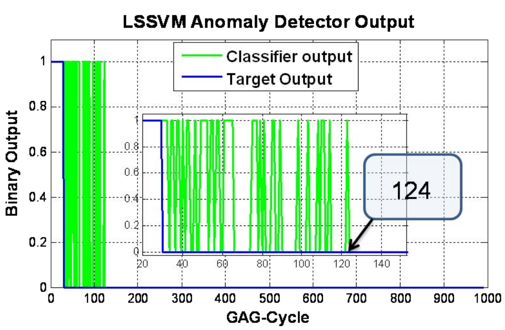

Figure 4-3 Feature correlated with evolving fault. 30 samples of initial feature data are used for training

When appended with the Bayesian Inference scheme, one notices that the detection capability is enhanced (GAG-cycle 115 with 95% confidence) and the results can be interpreted in terms of probabilities and confidence level.

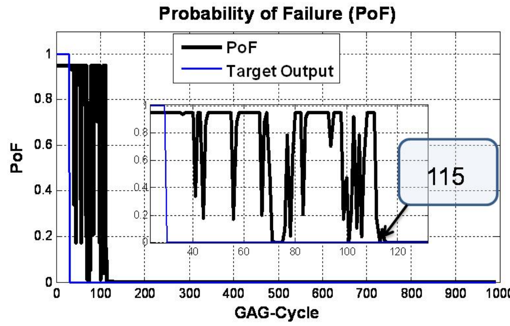

Figure 4-4 B-LSSVM:95% + confidence detection is possible at GAG Cycle 115

5. TEST CASE: DETECTION OF CRACKS IN BLADES OF A TURBINE ENGINE

Consider the case where the proposed methodology is applied to detect cracks in the blades of a turbine engine.

Figure 5-1 Picture of turbine engine under study

Light probes on both the leading and the trailing edge of the blades have been installed in order to provide with the Time-of-Arrival (TOA) for each blade. Since this is the only available piece of information in this application example, some pre-processing techniques were needed in order to generate a feature that can be used for detection purposes (Tangential Blade Position, TBP) (Marcos, 2003).

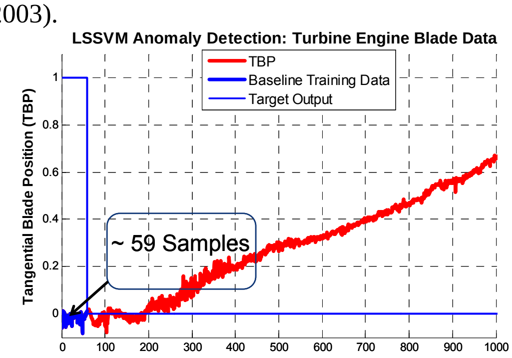

Figure 5-2 Feature correlated with evolving fault. 59 samples of initial feature data are used for training

The first few samples of the feature were used as Baseline training data as shown in Figure 5-2. Note that a sparse number of samples from the baseline are required in order to train the B-LSSVM Anomaly Detector. Results of the detection module for the case of one particular blade are shown in Figure 5-3 and Figure 5-4. Although it is possible to observe some changes in the probability of failure condition around the 200th cycle of operation, it is clear from both the classifier and the PoF plots that only after the 245th cycle that we can claim the existence of a fault with high confidence.

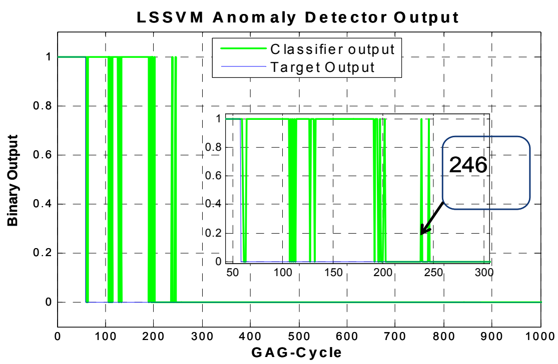

Figure 5-3 Feature correlated with evolving fault. 59 samples of initial feature data are used for training

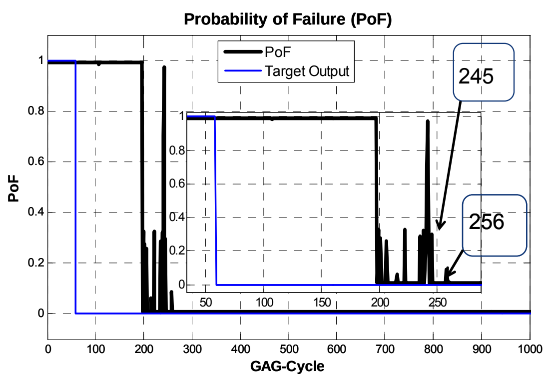

Figure 5-4 Feature correlated with evolving fault. 59 samples of initial feature data are used for training

6. CONCLUSION

A novel Anomaly Detector is suggested for one-class classification. The classifier is based on LSSVM machines suggested by Suykens and Vandewalle which they used for supervised binary learning. The proposed method requires only baseline data from the healthy system in order to train itself. Any novel phenomenon is then classified as an a*nomaly*. It has the additional advantage of having very low runtime overhead, the robustness to handle feature vectors as input and the ability to represent the diagnostics results probabilistically. Therefore, it is especially efficient in real-time diagnostics. Results from diagnosis of a crack on a Blackhawk aircraft gearplate and bending blades on the fan of a turbine generator are presented in this paper. Diagnostic results show encouraging advantages in using the one-class LS-SVM classifiers in terms of lowering the detection threshold while managing the region-of-uncertainty. This scheme is part of ongoing research at ICL labs. A more rigorous treatment of the proposed scheme including performance evaluation and comparisons with other contemporary Anomaly Detection approaches is in order and will be a part of future publications.

REFERENCES

Antoni, J. and R.B. Randall (2002), Differential diagnosis of gear and bearing faults. *Transactions of the ASME, Journal of Vibration and Acoustics, (124): p. 165-171*

Markou, M., Singh, S. (2003), Novelty Detection: A Review-part 1: Statistical Approaches, *Signal Processing, 83 (12): 2481-2497*

Markou, M., Singh, S. (2003), Novelty Detection: A Review-part 2: Neural Network based Approaches, *Signal Processing, 83 (12): 2499-2521*

Stefano, C. De, Sansone, C., Vento, M. ((2000)), To Reject or Not to Reject: That is The Questionan Answer in Case of Neural Classifiers, *IEEE Trans. Systems Man Cybernetics-Part C, 30 (1) 84-94*

Tax, M. J., Duin, P. W. (2004), Support Vector Data Description, *Machine Learning, 54: 45-66*

Dehmeshki, J., et al. (2004), Classification of Lung Data by Sampling and Support Vector Machine, in Engineering Medicine and Biology Society, *EMBC 2004. Conference Proceedings. 26th Annual International Conference of the, p. 3194 - 3197*

Soman, K.P., D.M. Shyam, and P. Madhavdas (2003), Efficient classification and analysis of ischemic heart disease using proximal support vector machines based decision trees, in TENCON, *Conference on Convergent Technologies for Asia-Pacific Region. 2003. p. 214-217*

Hu, Z., et al. (2005), Fusion of multi-class support vector machines for fault diagnosis, *in American Control Conference. Proceedings of the 2005, 1941 - 1945*

Schölkopf, B., Smola, A. J., Williamson, R. C., et al. (2000), “New Support Vector Algorithms”, *Neural Computation, vol. 5, pp.1207-1245*

Pan, M.-Q., et al. (2005), Support vector data description with model selection for condition monitoring, *in Machine Learning and Cybernetics, Proceedings of 2005 International Conference on. p. 4315 - 4318*

Suykens, J.A.K. and J. Vandewalle (1999), Least

Squares Support Vector Machine Classifiers, *Neural Processing Letters 9(3): p. 293-300*

Van Gestel T., Suykens J.A.K., Lanckriet G., Lambrechts A., De Moor B., Vandewalle J. (2002), “Bayesian Framework for Least Squares Support Vector Machine Classifiers, Gaussian Processes and Kernel Fisher Discriminant Analysis”, *Neural Computation, vol. 15, no. 5, pp. 1115-1148., Lirias number: 70607*

Boser, B.E., I.M. Guyon, and V.N. Vapnik (1992), A Training Algorithm for Optimal Classifiers, *in Proc. Fifth Ann. ACM Worksop Computational Learning Theory, D. Haussler, Editor.p. 144-152*

Bishop C.M. (1995), Neural networks for pattern recognition, *Oxford University Press*

MacKay D.J.C. (1992), “Bayesian Interpolation,” *Neural Computation, Vol.4, No.3, pp.415-447*

Van Gestel T., Suykens J.A.K., Lanckriet G., Lambrechts A., De Moor B., Vandewalle J. (2000), “A Bayesian Framework for Least Squares Support Vector Machine Classifiers,” *Internal Report 00-65, ESAT-SISTA, K.U.Leuven, submitted for publication.*

Chen, P.-H., C.-J. Lin, and B. Schölkopf (2005), A tutorial on v-support vector machines, *Applied Stochastic Models in Business and Industry, 21(2): p. 111-136.*

Wu, B., et al. (2005), Vibration monitoring for fault diagnosis of helicopter planetary gears, *International Federation of Automatic Control World Conference*

Zhang B., Khawaja T., Patrick R., Vachtsevanos G. (2007), Blind Deconvolution De-noising for Helicopter Vibration Data, *American Control Conference, ACC '07.*

APPENDIX A

The primal one-class LSSVM anomaly detector problem is defined according to Equation A.1:

$$\min_{w,\rho,e_k,\alpha_k} J_d = \frac{1}{2}\|w\| + \frac{\gamma}{2}\sum_{k=1}^{N} e_k^2 - \rho \qquad s.t.\ \langle w, \varphi(x_k)\rangle = \rho - e_k \tag{A.1}$$

Like all other SVM formulations an equivalent problem is constructed and solved used Lagrange multipliers as follows.

$$\mathcal{L}_d\left(w,\rho,e,\alpha\right) = J_d - \sum_{k=1}^{N}\alpha_k \left\langle w,\varphi(x_k)\right\rangle - \rho + e_k \tag{A.2}$$

where α_k are the Lagrange multipliers, which can be either positive or negative due to the equality constraints as follows from the Kuhn-Tucker conditions. The conditions of optimality lead us to the following expressions:

$$\begin{aligned}
\frac{d\mathcal{L}_d}{dw} &= 0 \Rightarrow w - \sum_{k=1}^{N}\alpha_k \varphi(x_k) = 0\\
\frac{d\mathcal{L}_d}{d\rho} &= 0 \Rightarrow \sum_{k=1}^{N}\alpha_k - 1 = 0\\
\frac{d\mathcal{L}_d}{de_k} &= 0 \Rightarrow \alpha_k - \gamma e_k = 0\\
\frac{d\mathcal{L}_d}{d\alpha_k} &= 0 \Rightarrow \sum_{k=1}^{N} w^T\varphi(x_k) - \rho + e_k = 0
\end{aligned} \tag{A.3}$$

for $k = 1,...,N$. These equations can be written in matrix form according Equation A.4.

$$\begin{bmatrix} I & 0 & 0 & -\varphi \\ 0 & 0 & 0 & I \\ 0 & 0 & \gamma I & -I \\ \varphi & -I & I & 0 \end{bmatrix}\begin{bmatrix} w \\ \rho \\ e \\ \bar{\alpha} \end{bmatrix} = \begin{bmatrix} 0 \\ \bar{1} \\ 0 \\ 0 \end{bmatrix} \tag{A.4}$$

This matrix formalization can be further simplified by eliminating and from Equation A.4 which leads us to the condensed Equation A.5 where $\Psi_{kl} = \varphi(x_k)^T\varphi(x_l)$.

$$\begin{bmatrix} 0 & \bar{1} \\ -\bar{1} & \psi + \gamma^{-1} I \end{bmatrix}\begin{bmatrix} \rho \\ \bar{\alpha} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \tag{A.5}$$

Taimoor S. Khawaja is a Graduate Research Assistant at Intelligent Control Systems Laboratory, Georgia Institute of Technology. He completed his Bachelors in Electrical Engineering from National University of Sciences and Technology, Pakistan and his Masters in Electrical and Computer Engineering from College of ECE, Georgia Institute of Technology, USA. His PhD thesis is based on exploring Artificial Intelligence techniques for data-driven Diagnostics and Prognostics in Condition Based Maintenance (CBM) systems. Currently, his research is focused on Anomaly Detection based on Bayesian Least Squares Support Vector Machines and real-term long-term prognostics via Bayesian Least Squares Support Vector Regression Machines. He has also designed and published algorithms for data denoising through Iterative Blind Deconvolution. He has been a part of several interesting projects related to diagnostics and prognostics of non-linear, time-varying Systems and vision based inspection and security systems.

Dr. George Vachtsevanos George Vachtsevanos is a Professor Emeritus of Electrical and Computer Engineering at the Georgia Institute of Technology and also serves as the chief scientist of Impact Technologies, LLC. He was awarded a B.E.E. degree from the City College of New York in 1962, a M.E.E. degree from New York University in 1963 and the Ph.D. degree in Electrical Engineering from the City University of New York in 1970. He directs the Intelligent Control Systems laboratory at Georgia Tech where faculty and students are conducting research in intelligent control, neuro-technology and cardio-technology, fault diagnosis and prognosis of large-scale dynamical systems and control technologies for Unmanned Aerial Vehicles. His work is funded by government agencies and industry. He has published over 250 technical papers and is a senior member of IEEE. Dr. Vachtsevanos was awarded the IEEE Control Systems Magazine Outstanding Paper Award for the years 2002-2003 (with L. Wills and B. Heck). He was also awarded the 2002-2003 Georgia Tech School of Electrical and Computer Engineering Distinguished Professor Award and the 2003-2004 Georgia Institute of Technology Outstanding Interdisciplinary Activities Award.