Combined Support Vector Novelty Detection for Multi-channel Combustion Data

Lei A. Clifton*, Hujun Yin*, David A. Clifton† and Yang Zhang‡
*School of Electrical and Electronic Engineering, University of Manchester
Email: l.wang-2@postgrad.manchester.ac.uk, h.yin@manchester.ac.uk
†Department of Engineering Science, University of Oxford
Email: davidc@robots.ox.ac.uk
‡School of Mechanical, Aerospace and Civil Engineering, University of Manchester
Email: y.zhang@manchester.ac.uk

***Abstract*— Multi-channel combustion data, consisting of gas pressure and two combustion chamber luminosity measurements, are investigated in the prediction of combustion instability. Wavelet analysis is used for feature extraction. A SVM approach is applied for novelty detection and the construction of a model of normal system operation. Novelty scores generated by classifiers from different channels are combined to give a final decision of data novelty. We compare four novelty score combination mechanisms, and illustrate their complementary relationship in assessing data novelty.**

I. Introduction

Combustion instability, caused by the resonant coupling between heat (generated by combustion) and acoustic pressure, is a major problem in the operation of jet engines and power generators. Early warning of combustion instability is required in order to prevent catastrophic failure. Combustion image data from a high-speed camera have been investigated to predict instability in [1], in which a Gaussian mixture model (GMM) was constructed to identify novel data. A Support Vector Machine (SVM) novelty detection method in [2] achieved earlier identification of combustion instability, and greater distinction between stable and unstable data classes than conventional GMM methods.

Two data sets D_1, D_2 generated by a Typhoon G30 combustor are investigated in this paper. The combustor was operated at atmospheric pressure, providing data in stable and deliberately unstable modes of combustion. Unstable combustion is achieved by increasing fuel flow rates above some threshold, with constant air flow rate. Both data sets represent measurements taken as the combustor moves from stable to unstable combustion modes.

Each data set consists of three channels (with sampling frequency of 1 KHz), the first of which is the gas pressure of the fuel methane (CH_4) in the main burner. For stable combustion, the swirl air flow rate was 0.039 Kgs^{-1}, the fuel supplied to the main and the pilot burners were fixed at flow rate 22.61×10^{-4} Kgs^{-1} and 10.20×10^{-4} Kgs^{-1} respectively. In order to initiate combustion instabilities, the flow rates of fuel supplied to the main and pilot burners were increased to 26.18×10^{-4} Kgs^{-1} and decreased to 4.37×10^{-4} Kgs^{-1} respectively. The second and third channels are luminosity measurements recorded within the combustion chamber. A bundle of fine optical fibres were mounted at the rear focal point of a Nikon 35 mm camera, such that all light passing through the front lens was collected. The flame luminosity from the combustion chamber was measured using this system. The fibre optic bundle was bifurcated, each channel connected to a photomultiplier (ORIEL model 70704). This allowed the measurement of chemiluminescent emitters of C_2 radicals (visible at light wavelength 513 nm), and the global intensity of unfiltered light, corresponding to the second and third channels in the data sets.

Support Vector Machines (SVMs) have been applied to many novelty detection problems, such as jet engine vibration testing [3], signal segmentation [4] and fMRI analysis [5]. The optical measurement methods described above have been applied to study flame dynamics of unstable combustion [6]. We apply one-class SVMs to similar multi-channel combustion data; classifier combination methods are then applied to fuse SVM classifications from different channels. This system is intended for on-line application, requiring efficient computation and minimal storage limitations. This paper compares two different fusion methods for novelty detection in Typhoon combustor operation.

II. Wavelet Feature Extraction

Frequency domain features alone are not suitable for on-line novelty detection in combustion data [1]. Features from wavelet analysis are used in the proposed method. Wavelet analysis represents a function in terms of basis functions, localised in both location and scale [7]. It is capable of revealing behavioural trends or transient periods within the data. Wavelet decomposition can be regarded as a multi-level or multi-resolution representation of a function $f(t)$, where each level of resolution j (except the initial level) consists of wavelets $\Psi_{j,k}$, with the same scale but differing locations k. Wavelet decomposition of a function $f(t)$ at level J can be written as

$$f(t) = \sum_k \lambda_{J,k} \Phi_{J,k}(t) + \sum_{j=J}^{+\infty} \sum_k \gamma_{j,k} \Psi_{j,k}(t), \qquad (1)$$

where $\Phi_{J,k}(t)$ are scaling functions at level J, and $\Psi_{j,k}(t)$ are wavelets functions at different levels j. $\lambda_{J,k}$ are scaling coefficients or approximation coefficients at level J. The set of $\gamma_{j,k}$ are wavelet coefficients or detail coefficients at different levels j.

Mallat [8] developed a filtering algorithm for Discrete Wavelet Transforms. Given a signal s, wavelet decomposition of the first level produces two sets of coefficients: approximation coefficients λ_1 and detail γ_1, by convolving s with a low-pass filter $h(k)$ for approximation, and with a high-pass filter $g(k)$ for detail, followed by dyadic decimation (downsampling). Wavelet decomposition of the next level splits the approximation coefficients λ_1 in two parts using the same scheme, replacing s by λ_1, and producing λ_2 and γ_2, and so on.

The energy e_j of the wavelet detail coefficients $\gamma_{j,k}$ within a window of data at level j reflects the average noise level of the signal within that window [9]. Coefficient energy within the window is defined to be

$$e_j = \frac{\sum_k \gamma_{j,k}^2}{L} \tag{2}$$

for window length L, and wavelet detail coefficients $\gamma_{j,k}$ at level j.

We investigate two triple-channel combustion data sets in this paper, which contain 5700 and 7400 data points respectively. These data are divided into 44 and 57 non-overlapping windows respectively, of length $L = 128$. We decompose each window of data using the Daubauchies-3 wavelet. The mean values of approximation coefficients λ_1 and the energy of the detail coefficients γ_1 (extracted at level $j = 1$) are used as two-dimensional features in the input space.

In previous work [2], the mean value of γ_1 was used; here, we extend the method to consider the energy of γ_1. We show in this investigation that use of such an energy-based feature results in earlier detection of non-normal combustion.

III. Support Vector Novelty Detection

Novelty detection, defined as detecting departures in behaviour from a model of system normality [10], is applicable to combustion systems in which often only "normal" data are available.

The training data available are l observations from "normal" operation $\mathbf{x}_1, \ldots, \mathbf{x}_l \in \mathbb{R}^d$. The training data can be mapped into another feature space $\mathbb{F}$ through a feature mapping $\mathbf{\Phi}$: $\mathbb{R}^d \rightarrow \mathbb{F}$. The kernel function operates on the dot product of the mapping function

$$k(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{\Phi}(\mathbf{x}_i) \cdot \mathbf{\Phi}(\mathbf{x}_j)). \tag{3}$$

A Gaussian kernel function is used here to suppress the growing distances for larger feature spaces [11]

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\sigma^2\right), \tag{4}$$

where σ is the width parameter associated with the kernel function.

The strategy developed in [12] maps the data into the feature space corresponding to the kernel function, and separates them from the origin with maximum margin. The decision function is found by minimising the weighted sum of a support vector type regulariser and an empirical error term depending on an overall margin variable ρ and individual errors ξ_i

$$\min_{w \in F, \xi \in \mathbb{R}^l, \rho \in \mathbb{R}} \quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l} \xi_i - \rho \tag{5}$$

$$\text{subject to} \quad (w \cdot \mathbf{\Phi}(\mathbf{x}_i)) \geq \rho - \xi_i, \quad \xi_i \geq 0, \tag{6}$$

where w is a weight vector in the feature space; C is a user-specified parameter, which reflects the boundary of normality region.

Using Lagrangian multipliers in the presence of constraints, and setting the derivatives of those multipliers with respect to w equal to zero, leads to

$$w = \sum_{i=1}^{l} (\alpha_i \mathbf{\Phi}(\mathbf{x}_i)) \tag{7}$$

$$\alpha_i = C - \beta_i \leq C, \quad \sum_{i=1}^{l} \alpha_i = 1, \tag{8}$$

where α_i and β_i $\{\ \alpha_i, \beta_i \geq 0\ \}$ are Lagrangian multipliers. In (7), patterns $\mathbf{x}_i$ with $\alpha_i > 0$ are the *support vectors*. The dual formulation is

$$\min_{\boldsymbol{\alpha} \in \mathbb{R}^l} \sum_{i,j} \alpha_i \alpha_j k(\mathbf{x}_i, \mathbf{x}_j) \tag{9}$$

$$\text{subject to} \quad \sum_i \alpha_i = 1, \quad 0 \leq \alpha_i \leq C. \tag{10}$$

Solutions α_i of the dual problem (9) determine parameters w_o and ρ_o:

$$w_o = \sum_{i=1}^{N_s} \alpha_i \mathbf{\Phi}(\mathbf{s}_i) \tag{11}$$

$$\rho_o = \frac{1}{N_s} \sum_{j=1}^{N_s} \sum_{i=1}^{N_s} \alpha_i k(\mathbf{s}_i, \mathbf{s}_j), \tag{12}$$

where N_s is the number of support vectors, and $\mathbf{s}_i$ are support vectors. Note that $w_o \in \mathbb{F}$; $\rho_o \in \mathbb{R}$.

The decision function is the hard limit function

$$f(\mathbf{x}) = \operatorname{sgn}\big(w_o \cdot \mathbf{\Phi}(\mathbf{x}) - \rho_0\big) \tag{13}$$

$$= \operatorname{sgn}\Big(\sum_{i=1}^{N_s} \alpha_i k(\mathbf{s}_i, \mathbf{x}) - \rho_0\Big). \tag{14}$$

For this investigation, we define the following function

$$g(\mathbf{x}) = \rho_0 - \sum_{i=1}^{N_s} \alpha_i k(\mathbf{s}_i, \mathbf{x}) \tag{15}$$

to assign novelty values to data, such that abnormal data (i.e. those outside the single, "normal" training class) take positive values between 0 and 1. The threshold above which data are deemed to be "abnormal" is zero, corresponding to the boundary of the decision function $g(\mathbf{x})$.

IV. Combination of Classifiers

Each window of data results in a two-dimensional feature vector $\mathbf{x}$, from which novelty values $g(\mathbf{x})$ are computed using (15). Resulting scores of features from each of the three data channels are combined. Each of the two data sets are considered independently.

Different classifier combination strategies have been studied in [13]–[15]. The mean combination rule has been used in previous analyses of combustion data [2]. In this paper, we compare the novelty detection capacity of the system using other methods of novelty score combination.

A. Max combination rule

We define a max combination rule to be

$$\hat{g}(\mathbf{x}) = \max_{i=1}^{R} g_i(\mathbf{x}), \tag{16}$$

B. Min combination rule

Similarly, we define a min combination rule to be

$$\hat{g}(\mathbf{x}) = \min_{i=1}^{R} g_i(\mathbf{x}), \tag{17}$$

C. Mean combination rule

Novelty scores generated by different classifiers may be combined according to the mean rule [14]:

$$\hat{g}(\mathbf{x}) = \frac{1}{R}\sum_{i=1}^{R} g_i(\mathbf{x}), \tag{18}$$

where $g_i(\mathbf{x})$ is the novelty score generated by the ith classifier, $i = 1, \ldots, R$, and R is the number of classifiers.

The max, min and mean rules are related by the following inequality

$$\min_{i=1}^{R} g_i(\mathbf{x}) \leq \frac{1}{R}\sum_{i=1}^{R} g_i(\mathbf{x}) \leq \max_{i=1}^{R} g_i(\mathbf{x}), \tag{19}$$

which suggests that the min and max rules serve as lower and upper bounds for the mean rule, respectively.

D. Product combination rule

For combination of novelty score $g(\mathbf{x})$, we here define a complimentary product combination rule to be

$$\hat{g}(\mathbf{x}) = \frac{\prod_{i=1}^{R} g_i(\mathbf{x})}{\prod_{i=1}^{R} g_i(\mathbf{x}) + \prod_{i=1}^{R}(1 - g_i(\mathbf{x}))}, \tag{20}$$

analogous to the combination rule for posterior probabilities in Bayesian classification [15]. This function reaches its upper bound at $\hat{g}(\mathbf{x}) = 1$ when feature vectors derived from windows of data are classified as maximally "abnormal" by all classifiers.

TABLE I

Data index of the first novel classification, according to each SVM classifier and combined efforts

	C^1	C^2	C^3	Max	Min	Mean	Product
D_1	18	19	18	18	23	18	24
D_2	23	28	23	23	23	23	30

V. Method, Results and Discussion

A. Method and Results

The series of wavelet features generated for the two data sets are two-dimensional for each channel, containing 44 and 57 windows respectively, and thus 44 and 57 feature vectors. The combustor operated in stable combustion mode until window index 20 in D_1, and index 28 in D_2. For D_1, this onset of non-normal behaviour occurs at elapsed time of $t \approx 2.7$ s, as shown in Fig. 1, in which the three channels of D_1 are plotted. Similarly, for D_2, the "normal" operation of the combustor ends at $t \approx 3.5$ s.

After this period of stable combustion, a period of transient behaviour is evident, prior to operation in unstable combustion mode for the remainder of both data sets. 80% of feature vectors from this stable period of operation (16 for D_1, 22 for D_2) are used as training data to construct the SVM classifiers. The remaining 20% data from stable operation, and all of the data from transient and unstable periods, are used as test data. A zero-mean, unit-variance transformation is applied to all data, for normalisation prior to the feature extraction [16].

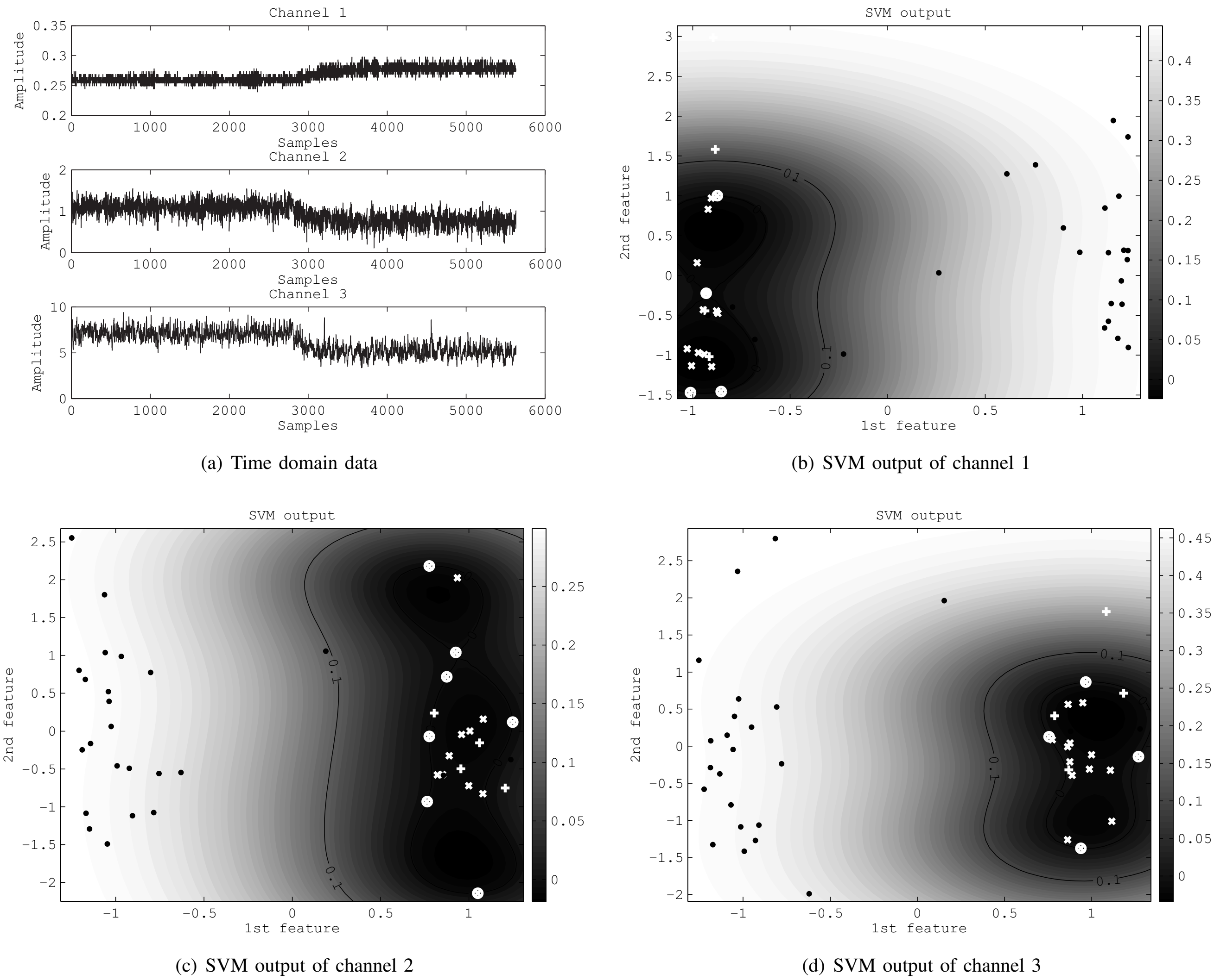
For each SVM classifier, we use empirically-determined parameter values for Gaussian kernel width $\sigma = 0.8$, and $C = 0.78$. The novelty threshold $H_s = 0$ is used, such that all data with novelty score $g_i(\mathbf{x}) > H_s$ are classified "abnormal" with respect to the one-class model, as determined by the ith classifier.

SVM classifiers are trained for each channel. Combining outputs from each classifier is performed using the proposed max, min, mean, and product rules. Contour plots of novelty scores by individual single-channel SVM classifiers are shown in Fig. 1. Novelty scores from combined SVM classifiers are shown in Fig. 2. Novel data are identified during transient periods by every classifier in all three channels of both data sets, providing confidence that non-normal combustor operation can be detected in this on-line context. The final decision on data abnormality is made by the combined classifier.

Table I shows the index of the first data point $\mathbf{x}$ for which $g_i(\mathbf{x}) > H_s$; i.e. when data are first deemed "abnormal" by each single-channel SVM classifier, $\{C^1, C^2, C^3\}$. The index of first "abnormal" data points according to the max, min, mean and product combination rules are shown as "Max", "Min", "Mean" and "Product", respectively.

B. Discussion

Single-channel SVM classifiers for both C^1 and C^2 show a peak in novelty scores at window indices 18 and 19, which



(a) Time domain data

(b) SVM output of channel 1

(c) SVM output of channel 2

(d) SVM output of channel 3

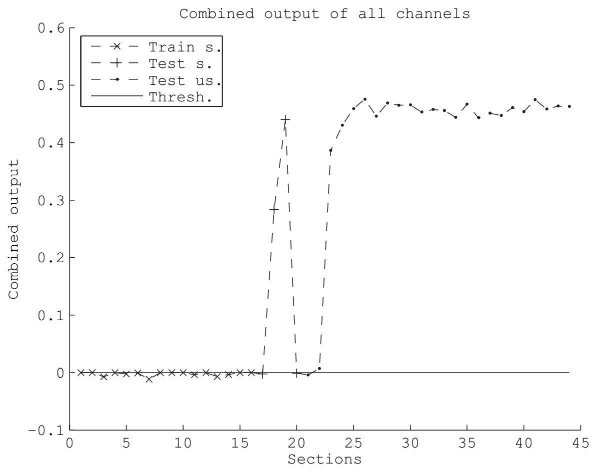
Fig. 1. Time-domain signals and individual SVM results for D_1. Stable training data, stable test data, transient test data, and unstable test data are shown by $\{\times + \cdot\}$ respectively. (a) Three channel data in time-domain: gas pressure of the main burner in the upper figure; C_2 radical luminosity in the middle figure, and global intensity of unfiltered light in the lower figure. (b)(c)(d) Contour plots of novelty scores generated by SVM classifers in the first, second and third channels respectively; support vectors are circled.

results in a corresponding peak value in combined novelty score with the *max* rule, shown in Fig. 2(a). In comparison, the single-channel classifier for C^3 gives a low output value, which is preserved by the *min* rule, in Fig. 2(b). It can be seen in Fig. 2(a) – (c) that the combined novelty scores produced by application of the max and min rules define the boundary of scores produced by the mean rule, as stated in (19).

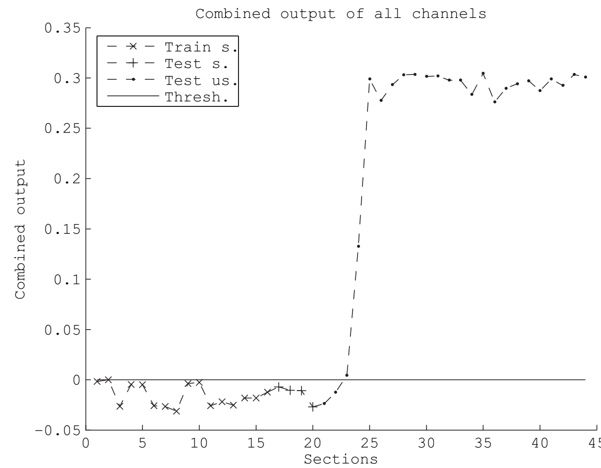
Similarly, results obtained using combining output from SVM classifiers with the *mean* rule show a clear peak at window indices 18 and 19 for D_1, prior to the onset of unstable combustion, as shown in Fig. 2. This is due to a similar peak in novelty scores output by the majority of individual single-channel SVM classifiers on C^1 and C^2, while there is no such a peak classifier on C^3. Thus, the peak present in two of the three channels is sufficiently significant to be retained after combination of novelty scores.

In comparison, no peak is evident in results from the combined classifier using the *product* rule. During windows 18 and 19, novelty scores of C^3 approach zero, which dominates the combined novelty score. However, as all three SVM classifiers show similar trends in novelty scores during later windows corresponding to the onset of unstable combustion, the combined classifier correctly identifies the unstable combustion. The product combination method attenuates novelty scores close to 0, as can be seen from the low variability in novelty scores associated with features derived from the "normal" operation of the system.

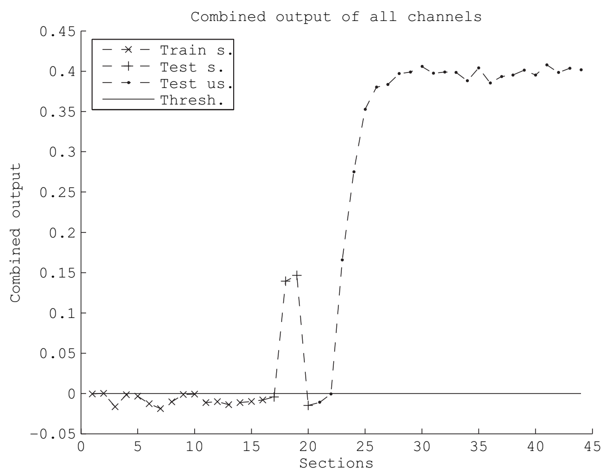
Similar results (not shown here) were obtained from comparison of the four combination methods using D_2, and in two other related data sets not described within this investigation. Again, combination with the mean rule provides identification of non-normal behaviour during the transient period of com-



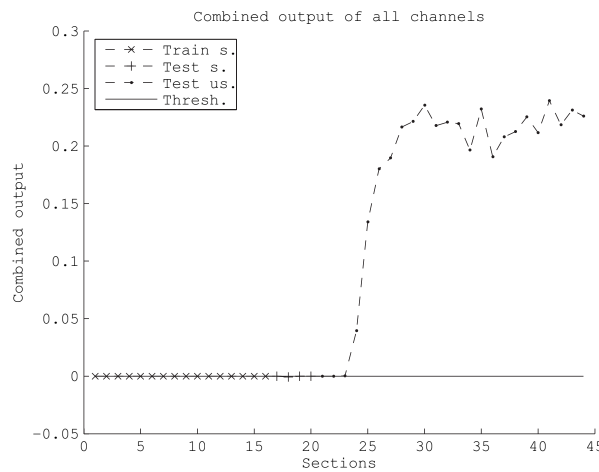
(a) Max combination rule



(b) Min combination rule



(c) Mean combination rule



(d) Product combination rule

Fig. 2. Novelty scores generated from the combined SVM classifier for D_1. Threshold $H_s = 0$ is marked by the horizontal line. Stable training data, stable test data, transient test data, and unstable test data are shown by $\{\times + \cdot\}$ respectively.

bustor operation, whilst exhibiting greater variation of novelty scores within the "normal" period of combustion. Results from using the min rule exhibit the variability in the "normal" region seen in results from use of the mean rule, whilst not providing early detection of the pre-cursor transient data. Results from using the max rule show that low variability in the "normal" region can be achieved while still providing detection of the pre-cursor transient behaviour.

VI. Conclusion

Combining classifier outputs using the mean rule has been shown to allow the detection of novelty earlier than combination with the product rule. Results have shown that precursors of unstable combustion may be detected using SVM classification combined using the mean rule, while the product rule provides lower variability in novelty scores during stable combustion, allowing a more sensitive novelty threshold H_s to be applied. The min combination rule provides little early warning of unstable combustion while suffering from variability in normal operation; the max combination rule provides optimal results, in which early detection of instability is possible without suffering significant variability in output during normal operation. A final decision on pattern abnormality can be made from observing the results of both combination rules.

Acknowledgment

The first author would like to thank members of the Advanced Flow Diagnostics and Instrumentation Laboratory for valuable assistance in running experiments for data collection.

References

[1] L. Wang and H. Yin, "Wavelet analysis in novelty detection for combustion image data," in *Proc. The 10th CACSC*, Liverpool, UK, 2004, pp. 79–82.

[2] L. Clifton, H. Yin, and Y. Zhang, "Support vector machine in novelty detection for multi-channel combustion data," in *Proc. ISNN (3)*, Chengdu, China, 2006, pp. 836–843.

[3] P. Hayton, L. Tarassenko, B. Scholkopf, and P. Anuzis, "Support vector novelty detection applied to jet engine vibration spectra," in *Proc. NIPS*, London, UK, 2000, pp. 946–952.

[4] A. Gretton and F. Desobry, "On-line one-class support vector machines. an application to signal segmentation." in *Proc. IEEE ICASSP*, Hong-Kong, China, 2003.

[5] D. R. Hardoon and L. M. Manevitz, "fmri analysis via one-class machine learning techniques," in *Proc. 19th International Joint Conference on Aritifical Intelligence (IJCAI)*, Edinburgh, UK, 2005, pp. 1604–1605.

[6] W. B. N. nad E. Clough, K. J. Syed, and Y. Zhang, "The combined investigation of the flame dynamics of an industrial gas turbine combustor using high-speed imaging and an optically integrated data collection method," *Measurement Science and Tachnology*, vol. 15, pp. 2303–2309, 2004.

[7] I. Daubechies, "Orthonormal bases of compactly supported wavelet," *Communications on Pure and Applied Mathematicse*, vol. 41, pp. 909–996, 1988.

[8] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.

[9] H. Guo, J. A. Crossman, Y. L. Murphey, and M. Coleman, "Automotive signal diagnostics using wavelets and machine learning," *IEEE Transactions on Vehicular Technology*, vol. 49, no. 5, pp. 1650–1662, 2000.

[10] D. A. Clifton, P. R. Bannister, and L. Tarassenko, "Learning shape for jet engine novelty detection," in *Proc. ISNN (3)*, Chengdu, China, 2006, pp. 828–835.

[11] D. M. J. Tax and R. P. W. Duin, "Data domain description using support vectors," in *Proc. ESAN99*, Brussels, 1999, pp. 251–256.

[12] B. Scholkopf, J. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Computation*, vol. 13, no. 7, pp. 1443–1471, 2001.

[13] J. Kittler, M. Hatef, R. Duin, and J. Matas, "On combining classifiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 226–239, 1998.

[14] L. I. Kuncheva, "A theoretical study on six classifier fusion stragies," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 2, pp. 281–286, 2002.

[15] D. M. J. Tax and R. P. W. Duin, "Combining one-class classifiers," in *Proc. Multiple Classifier Systems*, 2001, pp. 299–308.

[16] D. A. Clifton, "Condition monitoring of gas-turbine engines," University of Oxford, Tech. Rep., Jan. 2006.