Towards Real-Time and Memory Efficient Predictions of Valve States in Diesel Engines

Philippe Komma and Andreas Zell

***Abstract*— To reduce production costs of current engines, car manufacturers strive to replace built-in sensors by software solutions. However, the limitations of current micro controllers require time and memory efficient algorithms. In this paper, we propose a real-time framework for the detection of engine valve states based on wavelet analysis of in-cylinder pressure curves. Extracted wavelet features are then filtered out using mutual information such that only the most relevant wavelet coefficients become the input of the chosen support vector regressor. A further speedup is achieved by an approximation of the support vector solution which comprises less support vectors. We show that the combination of relevant feature selection and the regressor model simplification results in a significant decrease of the recall phase complexity while retaining good generalization performance.**

I. Introduction

In automotive engines, the system state can actively be changed via engine actuators. For example, valves that are part of the exhaust gas recirculation (EGR) scheme directly influence the redirection of emissions by establishing a pressure gradient. To this date, the state of these valves can only be determined using sensors, resulting in an increase of production costs. Thus, a software solution is preferred which predicts valve states using data of already existing sensors. Recently, the pressure sensor glow plug was introduced allowing for cylinder pressure onboard monitoring with regard to combustion control [1]. In this paper, we employ in-cylinder pressure curves acquired by this pressure sensor to detect the state of engine valves.

Several authors have addressed the usability of in-cylinder pressure in the domain of engine control. At first, research focused on establishing a model for in-cylinder pressure curves in various fields of applications [2], [3], [4]. As pointed out by Sellnau et al. [5], it does not only result in primary system benefits like an increased fuel economy or reduced NOx emmisions, but also in secondary effects like air-fuel balancing or calibration assistance.

With the advent of inexpensive cylinder pressure sensors [6], [7] new opportunities for precise engine control and fault detection have emerged. Park et al. [8] developed a spark advance control strategy based on the cylinder pressure signal in spark ignition engines. Therefore, they predicted the position of the peak pressure along with the hook-back at late burn-conditions using neural networks and employed this information to alter the spark advance. A fault diagnosis system for a diesel engine adapted from cylinder pressure was introduced by Yang [9]. In his work, he adopted a hybrid genetic algorithm and kernel principle component analysis technique for distinguishing engine valve errors from other machine faults.

In this paper, in-cylinder pressure curves are employed to predict the state of engine valves. Since all computations are performed on a micro controller of low clock rate, we do not only focus on the raw prediction performance, but also on the efficiency of the selected techniques. Here, efficiency is related to both run-time complexity and memory usage. In our approach, the reduction of the prediction complexity is based on two different strategies: first, we represent the signal by its most relevant features only, and second, the regression function is approximated to give a model of reduced complexity. As shown in Section V, this novel combined approach allows for real-time predictions of valve states on current micro controllers while maintaining good generalization performance.

The rest of this paper is organized as follows: after presenting our proposed method in Section II, we briefly introduce all employed preprocessing and regression techniques in Section III. Section IV comprises a detailed description of how these methods can be applied to the problem of valve state detection. Experimental results are presented and discussed in Section V and summarized in the last section.

II. Method Overview

The task of our proposed method is to estimate the state of two engine valves: the exhaust gas recirculation (EGR) valve embedded in the EGR system described in [10] and the variable turbine geometry (VTG) valve of a Garrett VTG system. While the former valve regulates the exhaust gas flow in the exhaust gas recirculation path, the latter valve effects the exhaust gas flow through the turbine. The term VTG valve comprises the VTG system consisting of the actuator, the mechanical drive component and the VTG vanes which are regulated rotatorily. Valve state changes are induced by modifying the control pulse which is represented by the duty cycle. The duty cycle is defined as the ratio between the pulse duration and the cycle duration. It is a dimensionless parameter measured in percent and ranges from 0% to 100%. Due to the more involved measurability of valve states in comparison with the duty cycle, we rather predict the latter. Note that for VTG valve prediction this approach introduces a potential error source since a well-defined relationship between the valve state and the exerted control pulse is not provided. For example, the exhaust pressure for a given duty

Ph. Komma and A. Zell are with the Chair of Computer Architecture, Computer Science Department, University of Tübingen, Sand 1, D-72076 Tübingen, Germany {philippe.komma, andreas.zell}@uni-tuebingen.de

cycle varies depending on whether the current state of the VTG vanes was reached by an opening or closing movement. Since the position of the VTG vanes influences the exhaust pressure this pressure difference indicates differing valve states for the same duty cycle. Valve state prediction issues may arise since the difference in the exhaust pressure is also reflected in the acquired in-cylinder pressure signal. Hysteresis tests for the EGR valve revealed a more distinct relationship between the valve state and the duty cycle.

Valve state estimation is based on pressure curves acquired in a single cylinder within a complete working cycle. We assume that all devices which are integrated in the air lines between the engine and the orifices influence the wave propagation in the gas exchange path. Hence, we expect a change of the pressure signal progression by modifying the valve settings. These changes are reflected in characteristic features contained in the acquired pressure curve. To extract these features the Haar wavelet transform is applied. Based on a mutual information filter approach we then choose the most relevant Haar wavelet coefficients. Wavelet coefficient selection is necessary to filter out those coefficients which represent the noise or other interferences contained in the pressure signal. The remaining coefficients constitute the inputs for both, the model training and the recall phase. Note that concentrating on the most important features only, involves the acceleration of our framework in two different ways: on the one hand, less wavelet coefficients have to be determined during the feature extraction step. On the other hand, valve state prediction complexity is decreased as well, since the complexity depends on the dimensionality of the input vector (see Section III-C).

In the model generation phase, the model learns the correct assignment of known valve states, given a selected set of Haar wavelet coefficients. In our approach, the model is represented by support vector machines (SVM). To meet real-time and memory constraints, we further simplify the established SVM by reducing the support vector set. In the recall phase, acquired and preprocessed pressure curves are applied to the simplified model in order to predict the state of engine valves.

III. Theoretical Background

A. Wavelet Analysis for Feature Extraction

Internal combustion engine acoustic measurements give information about the engine's operating parameters and physical characteristics. However, the acquired signals are complex and superimposed by backward noise, demanding accurate processing.

In our approach, we employ the Haar wavelet transform for feature extraction. As wavelets are localized in both space (time) and scale (frequency) domains, they can detect local features in a signal. Furthermore, wavelet analysis has a run-time complexity of $O(n)$ making it feasible in our domain. Mathematically, wavelet decomposition can be regarded as a multi-level representation of a function $f(t)$ that consists of a superposition of an approximation of itself at arbitrary scale index m_0 and a succession of signal details from scales m_0 down to $-\infty$:

$$f(t) = \sum_{n=-\infty}^{\infty} a_{m_{0,n}} \phi_{m_{0,n}}(t) + \sum_{m=-\infty}^{m_0} \sum_{n=-\infty}^{\infty} d_{m,n} \psi_{m,n}(t)$$

Approximation coefficients $a_{m,n}$ represent the signal at a coarser resolution, whereas wavelet coefficients $d_{m,n}$ describe the information lost when moving from an approximation of f at scale m to a coarser approximation at scale $m+1$. Wavelet functions $\psi_{m,n}(t)$ are generated from a single function ψ by dilations and translations:

$$\psi_{m,n} = |a|^{-\frac{1}{2}} \psi \left(\frac{t-b}{a} \right)$$

ψ is denoted as the mother wavelet function. By choosing a logarithmic discretization of parameters a and b, $\psi_{m,n} = \sqrt{2^{-m}} \psi \left((t - n2^m) 2^{-m} \right)$, a dyadic grid arrangement is established, forming the basis of the discrete wavelet transform (DWT). $\phi_{m,n}$ are denoted as scaling functions smoothing the signal. They have the same form as the wavelet:

$$\phi_{m,n(t)} = 2^{-m/2} \phi(2^{-mt-n}).$$

A fast wavelet decomposition scheme that computes the DWT using filters has been developed by Mallat [11]. It transforms a given signal s into two sets of coefficients: approximating coefficients a_1 and detail coefficients d_1 by convolving s with a low-pass filter and a high-pass filter. The outputs of both filters are then downsampled by a factor of two resulting in an output stream of approximation and detail coefficients that has the same length as s. This scheme is recursively repeated by applying the approximation coefficients as new inputs to the wavelet decomposition process.

B. Feature Selection using Mutual Information

Feature selection techniques are necessary for reducing the input dimensionality to avoid unwanted effects like overfitting. One solution for the feature selection process is to assign each feature a statistical relevance measure. If the chosen measure is independent from the employed regression model this technique is referred to as the filter or feature ranking approach [12].

Mutual information (MI) is a non-parametric measure of relevance which can be derived from information theory. The MI of two random variables x and y is a measure of how x and y depend on each other. It can be defined from the entropy $H(.)$:

$$\mathrm{MI}(x,y) = H(x) + H(y) - H(x,y) = H(y) - H(y|x),$$

where $H(y|x)$ is the conditional entropy of y given x. It measures the loss of uncertainty of y when x is known. If x and y are independent, then $H(x,y) = H(x) + H(y)$, $H(y|x) = H(y)$ and as a consequence $\mathrm{MI}(x,y) = 0$. For a continuous random variable x, mutual information

corresponds to the Kullback-Leibler distance between the joint distribution and the product of the marginals:

$$\begin{aligned} \mathrm{MI}(x,y) &= \mathrm{KL}(p(x,y)||p(x)p(y)) \\ &= \int\int p(x,y)\ln\left(\frac{p(x)p(y)}{p(x,y)}\right)dxdy \end{aligned}$$

In our work, we use the Kraskov MI estimator $I^{(2)}$ [13] which is based on entropy estimation using k-nearest neighbor statistics.

C. Support Vector Regression

In this section, we follow the results for support vector regression given by Smola and Schölkopf [14]. Given a training set $T = (x_i, y_i), i \in [1; l]$, where $x_i \in \mathbb{R}^d$, and $y_i \in \mathbb{R}$, we establish a linear regression function of the form:

$$f(x) = w^T\phi(x) + b \tag{1}$$

on a feature space F. Here, w denotes a vector in F and $\phi(x)$ maps the input x to a vector in F. w and b are obtained solving the following optimization problem:

$$\begin{aligned} \min_{w,b} P = \frac{1}{2}w^Tw + C\sum_{i=1}^{l}(\xi_i + \xi_i^*) \\ \text{s.t. } y_i - (w^T\phi(x+b)) \le \epsilon + \xi_i, \\ (x+b) - y_i \le \epsilon + \xi_i^*, \\ \xi_i, \xi_i^* \ge 0, i = 1, \ldots, l. \end{aligned}$$

The optimization criterion is chosen such that it penalizes data points for which $|y - f(x)| > \epsilon$. The slack variables, ξ and ξ^*, correspond to the lower and upper bounds in which the function $f(x) = w^T\phi(x)+b$ is allowed to deviate from a predefined error ϵ and cost C, where $\epsilon, C > 0$. The function $\phi(x)$ maps features x into a higher dimensional space.
By introducing Lagrange multipliers and formulating the corresponding Lagrangian the following dual optimization problem can be stated:

$$\begin{aligned} \min_{\alpha,\alpha^*} D &= \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}Q_{ij}(\alpha_i - \alpha_i^*)(\alpha_i - \alpha_i^*) + \\ &\quad \epsilon\sum_{i=1}^{l}(\alpha_i - \alpha_i^*) - \sum_{i=1}^{l}y_i(\alpha_i - \alpha_i^*) \end{aligned}$$

$$\begin{aligned} \text{s.t. } 0 \le \alpha_i, \alpha_i^* \le C, i = 1, \ldots, l, \\ \sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0, \end{aligned}$$

where $Q_{ij} = \phi(x_i)^T\phi(x_j) = K(x_i, x_j)$ and K denotes the kernel function. Given the solution of the dual optimization problem, the regression function of (1) can be written as:

$$f(x) = \sum_{i=1}^{l}\beta_i K(x, x_i) + b, \tag{2}$$

where $\beta_i = \alpha_i - \alpha_i^*$. It turns out that for a fraction of training points x_i, α_i and α_i^* equal 0. This results in a sparse solution since these training points can then be left out from (2) without altering the prediction result. The remaining n training points establish the predictive model and are denoted as support vectors. The simplified regression function becomes:

$$f(x) = \sum_{j=1}^{n}\widehat{\beta}_j K(x, \widehat{x}_j) + b. \tag{3}$$

Here, each support vector $\widehat{x}_j$ with weight $\widehat{\beta}_j$ correspond to a certain training point x_i for which either α_i or $\alpha_i^* \neq 0$.
As shown in Section V, the evaluation of (3) forms the limiting factor of the proposed valve state detection scheme. Since the prediction complexity is a function of support vector count n, the dimensionality d of x_i and the costs associated with the employed kernel function K, an increase in run-time performance involves the minimization of all three parameters. Whereas the theory of mutual information was introduced to optimize parameter d, the following section is aimed at reducing the number of support vectors.

D. Support Vector Reduction

Equation (3) reveals that a reduction of the support vector count results in both a decrease of run-time complexity in the recall phase as well as a decrease of memory requirements. The choice has to be made carefully since the rejection of even a small number of support vectors can result in a significant decrease in generalization performance as shown by Syed et al. [15]. In [16], Burges et al. reduced model complexity by introducing a *reduced set* of vectors. These vectors are generally not support vectors. In a later approach, Burges et al. [17] refined the method by determining the reduced set from the original vector set. Although promising, their method sometimes results in a convergence towards a local minimum as pointed out by Kwok and Tsang [18], rendering it necessary to restart the process with many initial guesses.
In our framework, we adopt the technique of Downs et al. [19]. In their work, they remove those support vectors that can be expressed as a linear combination of other ones in feature space. By modifying the SVM weights $\widehat{\beta}_i$, the generalization performance is preserved while the complexity of the regressor is decreased.
Let $\widehat{x}_i$, $i \in [1; n]$, be the set of support vectors. In the following, we assume that this set consists of r linear independent support vectors and $(n-r)$ support vectors that depend linearly on the other ones. Given that the ordering of the support vectors is chosen such that the first r support vectors are linearly independent, the regression function f of (3) can be formulated as:

$$f(x) = \sum_{i=1}^{r}\widehat{\beta}_i K(x, \widehat{x}_i) + \sum_{j=r+1}^{n}\widehat{\beta}_j\sum_{i=1}^{r}c_{ij}K(x, \widehat{x}_i) + b, \tag{4}$$

where K denotes the kernel function and M is the support vector matrix, $M \in \mathbb{R}^{l\times n}$, given by $M = K(\widehat{x}_i, x_j)$. Here, $\widehat{x}_i$ and x_j represent the set of n support vectors and the set of l training samples, respectively. The coefficient matrix C

describes how the linear dependent columns of the kernel matrix can be expressed using linear independent ones. It can be shown [19] that (4) is equivalent to:

$$f(x) = \sum_{i=1}^{r} \widetilde{\beta}_i K(x, \widehat{x}_i) + b, \tag{5}$$

where

$$\widetilde{\beta}_i = \widehat{\beta}_i + \sum_{j=r+1}^{n} \widehat{\beta}_j c_{ij}.$$

The application of the introduced support vector reduction scheme requires both, the determination of linear dependent vectors and the coefficient matrix $C = c_{ij}, \forall i, j$. Adopting the technique of [20] we therefore use QR factorization with column pivoting. It allows for the decomposition of a matrix M into an orthogonal matrix Q and an upper triangular matrix R, $M\Pi = QR$. Π is a permutation matrix, that sorts the columns of the matrix M according to the degree of their relative linear independence in decreasing order. Given that the rank of the kernel matrix M is r, the permuted kernel matrix $M\Pi$ can be decomposed into a set of r linear independent columns, M_1, and into another set, M_2, whose elements can be expressed as a linear combination of columns of M_1. Reformulating the QR factorization, we have:

$$\begin{bmatrix} M_1 & M_2 \end{bmatrix} = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{bmatrix}$$

The coefficient matrix C is then determined using the results of the QR decomposition:

$$M_1 C = M_2 \tag{6}$$

$$M_1 = Q_1 R_{11} \tag{7}$$

Solving (6) for C and inserting (7) into the resulting equation yields:

$$C = R_{11}^{-1} Q_1^T M_2.$$

The simplified regression function is an exact representation of (3) as long as the size of the reduced support vector set does not fall below the rank r of the support vector matrix M. In our experiments, the support vector count of the simplified models proved to be still too large to be applicable on a real engine controller, both in terms of run-time complexity and memory requirements. Hence, we did not consider the first r columns of the permuted support vector matrix $M\Pi$, but the first u columns instead. The reason of choosing this subset is to select those columns of the support vector matrix M which have the highest degree of linear independence and thus are most difficult to express by other columns. In this manner, (5) turns into an approximation of the exact solution as the value of u decreases. Note, however, that the rejected support vectors still influence the result of the approximated regression function as their weights $\widehat{\beta}_i, u < i \leq n$ are used to modify the weights $\widetilde{\beta}_i, 1 \leq i \leq u$ of the remaining support vectors.

IV. Experimental Setup

A. Data Acquisition

Although the state of engine valves is to be estimated from pressure curves, we performed data acquisition the reverse way. That is, given a set of operating points and valve positions the corresponding pressure curve taken from a single cylinder was sampled. Here, an operating point is determined by the number of revolutions per minute and the brake mean effective pressure value. The chosen set of operation points constituted an adequate representation of the complete engine map. Further, we considered three valve positions: a basis position that represented the standard position of an actuator and two positions that deviated from the standard one by 30% up and down. By varying both the valve setting and the operation point, we obtained a data set containing 500 observations. Given a fixed operating point and valve setting, the pressure curve of 100 working cycles was recorded after the engine reached its steady state. The sampling rate of the in-cylinder pressure sensor was set to two samples per degree crank angle. The value of the sampling rate was chosen to fulfill the Nyquist Theorem, i.e. not to lose any information due to undersampling. The sampling of a complete engine working cycle consisting of 720 degrees crank angle yielded 1440 samples representing the pressure progression.
Along with the pressure curve we measured three other engine parameters: the engine speed, the air flow value, and the amount of fuel added during the combustion process. Note that all three parameters were excluded from the following feature extraction and selection process, but directly used as inputs for the regressor.

B. Feature Extraction

For data preprocessing we implemented the DWT using the Haar basis, given by:

$$\psi(t) = \begin{cases} 1 & 0 \leq t < \frac{1}{2} \\ -1 & \frac{1}{2} \leq t < 1 \\ 0 & \text{elsewhere} \end{cases}$$

and

$$\phi(t) = \begin{cases} 1 & 0 \leq t < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Besides the three engine parameters mentioned above the feature vector representing a specific pressure curve consists of the concatenation of the approximation coefficient at scale $m = \lceil \log_2 1440 \rceil$ and wavelet coefficients of scales $m \in [1; \lceil \log_2 1440 \rceil]$. Note that this is only the preliminary feature vector since irrelevant coefficients are removed in the succeeding step. The removal of irrelevant features also contributes to the real-time applicability of the Haar wavelet transform (HWT). By formulating a modified version of the HWT which only determines the remaining coefficients we obtain a preprocessing scheme of low computational complexity. In comparison with the costs associated with valve state prediction the HWT run-time complexity proved to be insignificant, rendering it unnecessary to include the Haar wavelet transform in further complexity considerations.

C. Feature Selection

The aim of feature selection is twofold. At first, it partially solves the curse of dimensionality [21] by reducing the length of feature vectors significantly. Secondly, with regard to (3) the complexity of target value prediction is also a function of the feature vector length. This means, that by the sole selection of relevant feature components we do not only increase the possibility of obtaining a regressor model with a better generalization behavior, but we also increase the possibility of determining the results of the prediction faster. This is particularly important in this domain since the transformed input signal consists of 2048 features.

The easiest way of feature selection would be to perform an exhaustive search over all possible feature subsets. However, there are 2^{2048} subsets for this problem and therefore a brute force search is computationally infeasible. Hence, we used a filtering approach based on mutual information as described in Section III-B. Wavelet coefficients were ranked according to their mutual information with the target value in decreasing order. Based on this ranking the best d features were selected during the following model selection. In the next section, we compare this method against a low energy selection approach, where n components are selected, representing coefficients with lowest energy (largest scale). This choice is based on the assumption that the most relevant information is concentrated in the low energy content of the signal, whereas high energy coefficients only represent the signal's noise components.

D. Data Normalization

Given the matrix of all transformed pressure curves where each row represents a single observation, the columns were scaled to have a mean of zero and a standard deviation of 1. This avoids numerical problems during calculations and prevents wavelet coefficients of large magnitudes from dominating the training.

E. Model Selection

Model selection describes the process of determining the free parameters for the selected regressor. For the support vector machine this includes the specification of the kernel function K along with its parameters σ, the dimensionality d of the input space, the predefined error ϵ, and the cost value C. In our experiments, we used a Radial Basis Function Kernel $K(x, y) = \exp(-\|x - y\|^2 / 2\sigma)$, where x and y denote two feature vectors. Parameters σ, d, ϵ, and C were tuned to minimize the generalization error by a grid search. Each candidate parameter vector on the grid $(\sigma; d; C; \epsilon)$ was evaluated by a 10-fold cross-validation.

To accelerate the model selection phase, model training was performed using a reduced data set. We achieved data reduction by averaging the 100 pressure curves acquired for a fixed operating point and valve setting. This significantly reduced the time spent on model training whilst retaining good prediction results.

TABLE I

GENERALIZATION ERROR (MEAN ABSOLUTE ERROR, E_{mab}) OF (*a*) THE UNMODIFIED LOW ENERGY APPROACH AND (*b*) THE UNMODIFIED MUTUAL INFORMATION APPROACH. FOR EACH APPROACH THE RESPECTIVE SUPPORT VECTOR COUNT (*#SVs*), THE DIMENSIONALITY OF THE FEATURE VECTOR (*feat*), THE ESTIMATED NUMBER OF CLOCK CYCLES (*cycles*), AND MEMORY REQUIREMENTS (*mem*) ARE PRESENTED.

(a) unmodified low energy approach

valve	E_{mab}	feat	#SVs	cycles	mem (kB)
etvm	2.64	20	395	970269	32.4
vtg	5.60	24	357	932127	34.9

(b) unmodified mutual information approach

valve	E_{mab}	feat	#SVs	cycles	mem (kB)
etvm	1.70	52	375	1747022	77.6
vtg	4.92	32	343	1095246	44.2

F. Support Vector Reduction

For each feature set size $m \in [0; 64]$ with step size 4, we chose a support vector machine that provided a trade-off between generalization performance and run-time complexity. Therefore, we did not simply use the SVM with the best generalization performance, but the one with the smallest run-time complexity during valve state prediction which resulted in a decrease of generalization performance of less than 1%. In our investigations, the generalization performance is expressed in terms of the mean absolute error (E_{mab}). It is the absolute difference between the real and predicted target value, averaged over all test patterns and cross-validation folds. The estimation of the computational complexity is based on operation counting and assigning a certain cost factor to each operation. Since cost factors were not provided for the present automotive micro controllers we chose an alternative which resembled the reference most. In our experiments, run-time complexity analysis was performed on the basis of an Intel 386/87SX micro processor working at 25 MHz.

Finally, we applied the support vector reduction scheme (see Section III-D) to the selected support vector machines. By systematically reducing the support vector count given inputs of a certain dimensionality, we obtained an error function which yielded an estimate for the generalization performance depending on the chosen support vector count and feature set size. In a further step, we then chose the support vector machine which met defined accuracy or complexity constraints.

G. Target Value Prediction

Whereas the SVM training was performed using averaged pressure curves of 100 cycles, the prediction of target values was based on all recorded cycles. This gives rise to a more realistic testing procedure taking cyclic combustion variability, outliers, and noise into account. For measuring model quality the predicted target values were renormalized and the mean absolute error was determined over all training patterns.

TABLE II

ESTIMATED RUN-TIME COMPLEXITY MEASURED IN CLOCK CYCLES (CYLCES) RELATED TO THE GENERALIZATION ERROR (MEAN ABSOLUTE ERROR, E_{mab}) FOR THE ENGINE VALVE EGR USING THE LOW ENERGY (*le*) AND THE MUTUAL INFORMATION APPROACH (*mi*). FOR EACH GENERALIZATION ERROR BOUND THE PARAMETERS OF THE MINIMUM SUPPORT VECTOR MACHINE COMPRISING OF *#SVs* SUPPORT VECTORS OF DIMENSIONALITY *feat* ARE PRESENTED. ALL RELATIVE VALUES REFER TO THE RESULTS OF THE UNMODIFIED LOW ENERGY APPROACH (PRESENTED IN THE THIRD ROW) AND DENOTE THE PERCENTAGE OF REMOVED SUPPORT VECTORS AND CLOCK CYCLES, RESPECTIVELY.

E_{mab} inc		EGR (le)					EGR (mi)				
abs	rel	cycles	% rem	feat	#SVs	% rem	cycles	% rem	feat	#SVs	% rem
2.64	0	970269	0	20	395	0	970269	0	20	395	0
2.64	0	522900	46.1	20	249	40.5	86676	91.1	4	93	73.9
2.70	2.5	310780	68.0	12	205	51.0	82948	91.5	4	89	75.1
2.77	5	262268	73.0	12	173	58.7	82948	91.5	4	89	75.1
2.90	10	206856	78.7	8	169	59.6	75492	92.2	4	81	77.3
3.15	20	153000	84.2	8	125	70.1	68036	93.0	4	73	79.6
3.96	50	89352	90.8	8	73	82.6	45668	95.3	4	49	86.3

TABLE III

ESTIMATED RUN-TIME COMPLEXITY MEASURED IN CLOCK CYCLES (CYLCES) RELATED TO THE GENERALIZATION ERROR E_{mab} FOR THE ENGINE VALVE VTG USING THE LOW ENERGY (*le*) AND THE MUTUAL INFORMATION APPROACH (*mi*). FOR AN EXPLANATION OF THE EMPLOYED ABBREVIATIONS SEE TABLE II.

E_{mab} inc		VTG (le)					VTG (mi)				
abs	rel	cycles	% rem	feat	#SVs	% rem	cycles	% rem	feat	#SVs	% rem
5.60	0	932127	0	24	357	0	932127	0	24	357	0
5.60	0	547768	41.2	24	229	35.9	327248	64.9	16	181	49.3
5.74	2.5	286524	69.3	12	189	47.1	177372	81.0	12	117	67.2
5.88	5	256204	72.5	12	169	52.7	153116	83.6	12	101	71.7
6.16	10	167688	82.0	8	137	61.6	140988	84.9	12	93	73.9
6.69	20	123624	86.7	8	101	71.7	116732	87.5	12	77	78.4
8.40	50	41940	95.5	4	45	87.4	59976	93.6	8	49	86.3

V. RESULTS AND DISCUSSION

A. Valve State Prediction Performance

We begin our discussion by comparing the best results obtained after applying the proposed feature and model selection schemes without further simplification of the generated SVM models. The approach which employs mutual information filtered wavelet coefficients as input is denoted as the unmodified mutual information approach in the following. Using the low energy wavelet coefficients only, this approach is referred to as the unmodified low energy approach. Table I shows that support vector machines are capable of predicting the state of engine valves. Thereby, the mutual information feature selection approach results in an increase of prediction performance as compared to the low energy approach (55.3% for the EGR valve and 13% for the VTG valve). Note, however, that both feature selection approaches generate models of high computational costs rendering it impossible to achieve the required prediction frequency of 50 Hz. This statement particularly holds for the mutual information approach where more features are necessary to establish the model with the largest generalization performance.

B. Prediction under Complexity Constraints

Tables II and III show the results when establishing a trade-off between generalization performance and prediction complexity. Given a reference mean absolute error we searched for the reduced SVM model of lowest complexity that deviated from the reference by a predefined percentage (E_{mab} inc,rel). As reference we chose the mean absolute error of the unmodified low energy approach. The prediction complexity was estimated using the procedure introduced in Section IV-F taking both the number of features and the number of support vectors into account. Beside presenting the absolute values for the estimated cycle count and the support vector count, the relative decrease of both values is given (% rem). The decrease denotes the percentage of removed support vectors and clock cycles, respectively. Again, we use the results of the low energy approach as reference. We further differentiate between the reduced low energy (le) and the reduced mutual information (mi) approach representing support vector machines of reduced support vector set size. For the reduced low energy feature selection approach, the applied support vector reduction technique already removes 46.1% of all support vectors without losing any generalization performance. Yet, the number of employed features is still large resulting in a significant prediction complexity

TABLE IV

THE ABSOLUTE GENERALIZATION ERROR (MEAN ABSOLUTE ERROR, E_{mab}) AND THE RELATIVE INCREASE IN GENERALIZATION ERROR (*inc*) RELATED TO CALCULATION CAPACITY (*usage*) WHICH IS AVAILABLE FOR EGR VALVE STATE PREDICTION (*left*) AND VTG VALVE STATE PREDICTION (*right*). ALL RELATIVE VALUES REFER TO THE RESPECTIVE MEAN ABSOLUTE ERROR OF THE UNMODIFIED LOW ENERGY APPROACH (PRESENTED IN THE THIRD ROW). NEGATIVE VALUES INDICATE AN INCREASE IN GENERALIZATION PERFORMANCE.

	EGR (le)		EGR (mi)	
usage (%)	E_{mab} (%)	inc (%)	E_{mab} (%)	inc (%)
194	2.64	0	2.64	0
100	2.66	0.87	1.80	-31.85
50	2.81	6.62	1.90	-27.98
30	3.18	20.47	1.97	-25.15
20	3.80	43.93	2.33	-11.58

	VTG (le)		VTG (mi)	
usage (%)	E_{mab} (%)	inc (%)	E_{mab} (%)	inc (%)
186	5.60	0	5.60	0
100	5.67	1.18	5.32	-5.10
50	5.91	5.57	5.61	0.22
30	6.27	12.00	5.89	5.16
20	6.72	19.99	7.11	26.87

TABLE V

THE ABSOLUTE GENERALIZATION ERROR (E_{mab}) AND THE RELATIVE INCREASE IN GENERALIZATION ERROR (*inc*) RELATED TO THE MEMORY (*mem*) REQUIRED FOR STORING THE SVM MODEL WHEN PREDICTING EGR VALVE STATES (*left*) AND VTG VALVE STATES (*right*). ALL RELATIVE VALUES REFER TO THE RESPECTIVE MEAN ABSOLUTE ERROR OF THE UNMODIFIED LOW ENERGY APPROACH (PRESENTED IN THE THIRD ROW).

	EGR (le)		EGR (mi)	
mem (kB)	E_{mab} (%)	inc (%)	E_{mab} (%)	inc (%)
32.4	2.64	0	2.64	0
32	2.65	0.38	1.70	-35.54
16	2.70	2.22	1.82	-30.89
8	2.80	6.32	1.90	-27.99
4	3.21	21.69	1.91	-27.52
2	4.28	62.43	2.36	-10.43

	VTG (le)		VTG (mi)	
mem (kB)	E_{mab} (%)	inc (%)	E_{mab} (%)	inc (%)
34.9	5.60	0	5.60	0
32	5.65	0.83	4.92	-12.18
16	5.67	1.18	5.44	-2.92
8	5.91	5.44	5.61	0.22
4	6.14	9.66	6.32	12.75
2	6.72	19.99	7.38	31.68

in the recall phase. As for the reduced mutual information approach, we obtain a reduction of the estimated cycle count by more than 91% while maintaining the same generalization performance as compared to the unmodified low energy approach. The decrease in complexity arises from the decline of both the number of employed features and the number of support vectors. Already a set size of 4 features suffices to adequately predict the EGR valve state. Further improvements of the computational complexity at the expense of a slight increase of the generalization error are achieved by the reduction of the support vector count. A reduced support vector machine which is composed of more than 95% less support vectors as compared to the unmodified low energy feature selection approach yields absolute prediction errors of approximately 4%.

For the VTG valve we examine a similar behavior of the computational complexity in relation to a chosen generalization error bound if the generalization performance of the unmodified low energy approach should be maintained. Here, the reduced mutual information approach requires a smaller feature set as well as less support vectors as compared to the reduced low energy approach. This results in a decrease of computational complexity by 40.3%. If we allow for a larger generalization error, however, the difference in complexity becomes smaller and at a relative increase of the mean absolute error by 50% the reduced low energy approach is faster than the reduced mutual information approach. Table III indicates that the reduced complexity of the reduced low energy approach is attributed to the smaller dimensionality of the feature vector. In the case of VTG valve state prediction, the mutual information feature selection scheme identifies relevant features which yield improved generalization performance when they are applied as a whole, yet the generalization performance decreases if the chosen subset of relevant features becomes too small.

C. CPU Load Considerations

We expect that only a fraction of the complete processing power of the micro controller is available for valve state prediction. Based on our complexity estimates, Table IV presents the generalization errors which are obtained for a given CPU load. In this investigation, we assume a micro controller working at 25 MHz.

For the EGR valve, the reduced mutual information approach achieves good generalization performance even at high CPU loads. At a CPU load of 80% the mean absolute error is still more than 11% smaller as compared to the unmodified low energy approach. In this case, the mutual information approach also outperforms the reduced low energy approach by more than 63%.

If the available calculating capacity is higher than 20% VTG valve state prediction benefits from mutual information feature selection. Only at a CPU load of 80% the reduced low energy approach outperforms the reduced mutual information approach. As stated in the previous section the decreased performance of the reduced mutual information approach is primarily based on the enlarged feature set size related to the reduced low energy approach.

D. Memory Requirements

From Table V it can be derived that valve state prediction based on unmodified support vector machines involves a memory usage of $34.9\,\text{kB}$-$77.6\,\text{kB}$ per valve. Since micro controllers are memory bounded to a great extent we now consider reduced support vector machines of low memory requirements only. Note that all tests have been performed using $32\,\text{bit}$ floating point numbers. This yields memory requirements of $4(d{\cdot}n{+}n{+}1)$ bytes in size, where d denotes the dimensionality of a support vector and n is the support vector count.

As regards to the EGR valve a reduced support vector machine occupying $2\,\text{kB}$ in memory yields best generalization performance if the mutual feature selection procedure is used. In comparison with the unmodified low energy approach we obtain a decrease in memory requirements of more than 94%. For the VTG valve, strict memory constraints result in a loss of generalization performance. Given a memory of $4\,\text{kB}$ the mean absolute error rises by 9.7% and 12.75% for the reduced low energy approach and the reduced mutual information approach, respectively. Again, the generalization performance of the unmodified low energy approach is defined as the reference.

VI. Conclusion

In this paper, we showed that engine valve states can be predicted by in-cylinder pressure curves along with three additional engine parameters. To enable real-time predictions even on limited micro controllers we adopted a framework based on an effective preprocessing strategy using accelerated Haar wavelet analysis, a mutual information relevance measure that filtered out irrelevant wavelet coefficients, and a simplified support vector solution. In comparison with the low energy feature selection technique, which only selects wavelet coefficients of low energy, the mutual information feature selection yielded the best generalization performance given no memory and run-time constraints. For the EGR valve this finding also holds true if memory or the calculating capacity is limited. As for the VTG valve, the low energy feature selection approach resulted in a more compact solution in the case of a considerably constrained setting. In combination with the presented support vector reduction scheme the complexity of the recall phase was decreased by an order of a magnitude without significantly effecting the generalization performance. The obtained prediction results for both valves leave us optimistic that our valve state detection framework using in-cylinder pressure curves can be adopted for control purposes on automotive micro controllers.

In further work, we will investigate the robustness of our prediction framework with regard to various factors of influence. In this context, we will address issues arising from engine transients, production-based vehicle-to-vehicle varieties and engine aging. Later research will also focus on an enlarged acceleration of the SVM recall phase. Using optimization algorithms we will identify the set of support vectors which allows for a tighter approximation of the support vector regression function.

References

[1] J. Hadler, F. Rudolph, R. Dorenkamp, M. Ksters, D. Mannigel, and B. Veldten. Volkswagen's new 2.0 l TDI engine for the most stringent emission standards. *MTZ Worldwide*, 69(06), 2008.

[2] A. Shidfar and M. Garshasbi. Estimation of cycle-resolved in-cylinder pressure and air-fuel ratio using spark plug ionization current sensing. *Applied Mathematics and Computation*, 165:163–170, June 2005.

[3] H. Hamedovic, F. Raichle, and J.F. Bohme. In-cylinder pressure reconstruction for multicylinder SI-engine by combined processing of engine speed and one cylinder pressure. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05)*, volume 5, pages 677–680, Mar. 2005.

[4] D. Brand, L. Guzzella, and C. Onder. Estimation of the instantaneous in-cylinder pressure for control purposes using crankshaft angular velocity. *SAE transactions*, 114:426–437, 2005.

[5] M.C. Sellnau, F.A. Matekunas, P.A. Battiston, C.F. Chang, and Lancaster. Cylinder-pressure-based engine control using pressure-ratio management and low-cost non-intrusive cylinder pressure sensors. SAE Paper No. 2000-01-0932, Mar. 2000.

[6] M. Wlodarczyk. High accuracy glow plug-integrated cylinder pressure sensor for closed loop engine control. SAE Paper No. 2006-01-0184, Apr. 2006.

[7] T.J. Poorman, S. Kalashnikov, M.T. Wlodarczyk, and E. Danielson. In-cylinder, fiber-optic pressure sensor for engine monitoring and control. SAE Paper No. 981913, Aug. 1998.

[8] S. Park, P. Yoon, and M. Sunwoo. Cylinder pressure-based spark advance control for SI engines. *JSME International Journal. Series B, Fluids and Thermal Engineering*, 44(2):305–312, 2001.

[9] W.X. Yang. Vibration-based diagnosis of engine valve faults. *Insight*, 45:547–553, 2003.

[10] W. Wenzel, V. Joergl, P. Keller, and O. Weber. EGR systems for diesel engines - the road ahead. In *SIA International Conference - DIESEL ENGINES: The low CO2 and Emissions Reduction Challenge*, 2008.

[11] S.G. Mallat. A theory for multiresolution signal decomposition - the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:674–693, July 1989.

[12] I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3:1157–1182, 2003.

[13] A. Kraskov, H. Stögbauer, and P. Grassberger. Estimating mutual information. *Physical Review E*, 69(6):066138, June 2004.

[14] A.J. Smola and B. Schölkopf. A tutorial on support vector regression. *Statistics and Computing*, 14(3):199–222, 2004.

[15] N.A. Syed, H. Liu, and K.K. Sung. Incremental learning with support vector machines. Workshop on SV Machines at International Joint Conference on Artificial Intelligence (IJCAI99), Stockholm, 1999.

[16] C.J. Burges. Simplified support vector decision rules. In *Proceedings of the 13th International Conference on Machine Learning*, pages 71–77. San Mateo, CA, Morgan Kaufmann, 1996.

[17] C.J. Burges and B. Schölkopf. Improving speed and accuracy of support vector learning machines. *Advances in neural information processing systems*, 9:375–381, 1997.

[18] J.T.-Y. Kwok and I.W.-H. Tsang. The pre-image problem in kernel methods. *IEEE Transactions on Neural Networks*, 15(6):1517–1525, Nov. 2004.

[19] T. Downs, K.E. Gates, and A. Masters. Exact simplification of support vector solutions. *Journal of Machine Learning Research*, 2:293–297, 2002.

[20] A. Pronobis and B. Caputo. The more you learn, the less you store: Memory-controlled incremental SVM. In *2nd International Cognitive Vision Workshop (ICVW06)*, Graz, Austria, May 2006.

[21] P.C. Hammer. Adaptive control processes: A guided tour. *SIAM Review*, 4(2):163–163, 1962.