ABSTRACT

In this paper, a framework for isolating unprecedented faults for an EGR valve system is presented. Using normal behavior data generated by a high fidelity engine simulation, the recently introduced Growing Structure Multiple Model System (GSMMS) is used to construct models of normal behavior for an EGR valve system and its various subsystems. Using the GSMMS models as a foundation, anomalous behavior of the entire system is then detected as statistically significant departures of the most recent modeling residuals from the modeling residuals during normal behavior. By reconnecting anomaly detectors to the constituent subsystems, the anomaly can be isolated without the need for prior training using faulty data. Furthermore, faults that were previously encountered (and modeled) are recognized using the same approach as the anomaly detectors.

1 INTRODUCTION

In order to meet the increasingly demanding emission and fuel consumption requirements, automotive systems have become increasingly complex and sophisticated. The push to meet emissions standards has given rise to the use of many controlled systems that must be monitored in order to ensure performance. One such system is the Exhaust Gas Recirculation (EGR) system which is used to lower combustion temperatures and reduce NOx emissions [1]. Since EGR has become an integral part of emission control, ensuring proper performance is essential to maintaining low emissions and high performance.

To this end, on-board diagnostic (OBD) systems have been developed that follow the traditional precedent-based diagnostic paradigm. OBD systems run through a series of system checks looking for a set of predetermined faulty conditions. These faults have been anticipated by designers or have been observed in the past. System signatures extracted from sensor readings in the presence of those faults are then used to train models (statistical, dynamic, rule-based) of faulty system behavior, based on which the presence of the particular fault can be recognized by matching models of current system behavior with known fault models. However, as automotive systems become increasingly complex, it becomes infeasible to train for all possible faults, under all possible driving conditions. Clearly, it is advantageous to create a scheme for isolation of the anomalous subsystem that requires little a priori knowledge and does not require faulty training data.

In this paper, the method for precedent-free fault detection and isolation introduced in [2] is applied to an EGR valve system. The methodology is comprised of anomaly detection, fault isolation and fault diagnosis. Any intrusion or degradation in system performance is identified as a statistically significant departure from a model of normal system behavior. Subsequently, if an anomaly is found, multiple anomaly detectors connect to the relevant subsystems of the anomalous system, with each detector that detects anomalous behavior splitting further into anomaly detectors monitoring subsystems of ever increasing granularity. It has been shown that such multiplication of anomaly detectors leads to fault isolation, even if the fault has never been observed [3].

Anomaly detection and isolation is followed by fault diagnosis. This is accomplished by matching the input-output patterns of the anomalous subsystem with models of previously seen faults. If the observed patterns cannot be matched by any of the existing fault models, a new model is then created in order to recognize that particular situation in the future.

The aforementioned method was successfully demonstrated in precedent-free anomaly isolation in an electronic throttle system (ETS) [2]. In this paper, a precedent-free anomaly detection method is applied to a diesel engine EGR valve system, which
can be seen as a significantly more challenging diagnostic problem. Namely, prominent non-linearities are present in every constituent subsystem of an EGR valve system, which imposes challenges in terms of dynamic system modeling, anomaly detection and fault diagnosis. Furthermore, it involves a larger number of constituent subsystems compared to the ETS (only 2 subsystems considered in [2], only one of which was non-linear), which further complicates anomaly isolation and diagnosis.

The remainder of the paper is organized as follows. Section 2 describes the anomaly detection procedure that utilizes the data-driven Growing Multiple Model System (GSMMS) approach to modeling dynamic systems. Section 3 contains the results of the application of this methodology to the EGR system on a diesel engine. Finally, Section 4 states the conclusions and suggests future work.

2 GSMMS BASED ANOMALY DETECTION, ISOLATION AND DIAGNOSIS

The key enabling device for the method proposed in [2] is a generic modeling approach, allowing anomaly detectors and diagnosters to use essentially the same modeling mechanism to successively connect to different inputs and outputs (corresponding to different subsystems), identify models of the corresponding subsystems and accomplish anomaly detection and diagnosis. The modeling approach should therefore possess sufficient analytical tractability to enable modeling of system dynamics as well as anomaly detection and diagnosis with mathematical rigor and statistical significance. The Growing Structure Multiple Model System (GSMMS) [2] approach to dynamic system modeling provides such a generic modeling tool with locally simple models that enable the precedent-free methodology utilized in this paper.

The following sections will discuss the anomaly detection, isolation and diagnosis procedure in further detail. Section 2.1 will discuss the GSMMS and its structure, Section 2.2 will discuss the training procedure, Section 2.3 will detail the interpretation of the residuals and the determination of a fault model.

2.1 GSMMS Modeling Approach

Various data driven modeling methods have been used for identification of complex dynamic systems. They impose fewer assumptions on the underlying model while requiring less a priori knowledge compared to the first principle physics based models. Methods such as Multi-Layer Perceptrons (MLPs) and Recurrent Neural Networks (RNNs) have been used extensively for data driven modeling due to their universal approximation capabilities [4]. However, the need for exhaustive and elaborate training, problems with generalization outside the training set and lack of analytical tractability of the resulting models tend to limit their practical utility.

An alternative to such “global” modeling methods is the “divide and conquer” approach where the operating space is decomposed into smaller sub-regions. A variety of frameworks have been proposed for such multiple model systems. Takagi and Sugeno [5] proposed a model that creates a fuzzy partition of the input space using membership functions. Although the Takagi-Sugeno Fuzzy Model (TSFM) can apply well known methods (such as the Kalman Filter) to estimate the local model parameters, the model structure (premise variables) must be determined using a heuristic. The problem of determining both local parameters and model structure is addressed by NeuroFAST [6] which utilizes fuzzy adaptive resonance theory along with fuzzy rule splitting and addition to provide better coverage of regions that are difficult to model. An approach proposed by Johansen and Foss [7] computes the global output using the following structure:

$$ y = \sum_{i=1}^{M} v_i(x) f_i(x) $$

$$ v_i(x) = \frac{p_i(x)}{\sum_{i=1}^{M} p_i(x)} $$

where $f_i(x)$ is the local model, $y$ is the output and $p_i(x)$ describes the validity of the $i^{th}$ model. The global output can be seen as a weighted sum of the local models. However, the regime decomposition relies on heuristics and is limited to a rectangular grid. For this model and the TSFM, the problem of determining the structure is computationally complex.

Instead of dividing the individual input variables, vector quantization techniques such as a Kohonen’s Self-Organizing Map (SOM) [8], have been proposed to identify a dynamic model structure [9] [10]. The set of weight vectors $\xi_i, i = 1, \ldots, M$ of a SOM define a Voronoi Tessellation:

$$ V_m = \{ x : \| x - \xi_m \| \leq \| x - \xi_j \|, \forall j \neq m \} $$

If a SOM is used to cluster vectors consisting of inputs and outputs of a dynamic system, the Voronoi Tessellation ends up partitioning the input-output space into sub-regions of “similar” input-output patterns [11]. An advantage of this method is that the partition is adjusted through unsupervised clustering carried out by modifications of the SOM. In this respect, the SOM acts as a set of “decoders” to order the space [8]. However, the number of nodes and the number of topological connections still need to be determined in advance.

Recent developments in growing SOMs allow the maps to grow to appropriate size and impose fewer assumptions on the underlying data. Growing neural gas [12], growing cell structures [13] and growing SOMs [14] incorporate specific addition and deletion mechanisms to allow the SOM to determine their appropriate order and quantity.

These developments led to the recently introduced Growing Structure Multiple Model System [2] approach to dynamic system modeling. It utilizes the growing SOM to decompose the
input space into sub-regions with similar dynamic behavior and to refine the partition in poorly modeled regions.

More specifically, the GSMMS uses the multiple models structure of Eq. 1 where each of the $M$ regions are defined by the SOM induced Voronoi Tessellation. Local models are assumed to be of the linear form

$$F_m(s(k)) = a_m^T s(k) + b_m$$

(3)

with $a_m$ and $b_m$ denoting the parameters of the local model $m$ and

$$s(k) = [y^T(k), \ldots, y^T(k-n_a+1), u^T(k-n_d), \ldots, u^T(k-n_d-n_b+1)]^T$$

(4)

where $y(k) = [y_1(k), \ldots, y_s(k)]^T$ and $u(k) = [u_1(k), \ldots, u_p(k)]^T$ for a system with $n$ outputs and $p$ inputs. As in Johansen and Foss [7], the global model is then defined as an interpolation of local models

$$\hat{y}(k + 1) = \sum_{m=1}^{M} v_m(s(k)) F_m(s(k))$$

(5)

where $v_m(s(k))$ is the function describing the way local models are mixed into the global model. Following Liu [2], $v(s(k))$ is taken to be a simple gating function

$$v_m(s(k)) = \begin{cases} 1 & s(k) \in V_m \\ 0 & \text{otherwise} \end{cases}$$

(6)

which defines the region of validity of each local model as one of the Voronoi sets of the underlying SOM. Subsequently, this structure implies that only one local model is used to describe the dynamics at a given sampling instant, $k$.

One can see that the GSMMS essentially casts the problem of representing the system dynamics into the framework of simple, interconnected dynamic linear models. This structure enables the modeling of a wide variety of complex systems while maintaining analytical tractability. Local model tractability recently led to important results in terms of global characteristics of the model parameter estimation during learning [15], model stability [16] and is expected to lead to describable control capabilities [3]. The GSMMS has been used successfully for modeling of an electronic throttle system in a gasoline engine [2] and automotive crankshaft dynamics [17].

### 2.2 Training The GSMMS Model

In order to model a system using the GSMMS, both the structural parameters of the model (weight vectors) and local model parameters must be determined. The state space partition is updated by adjusting the SOM weight vectors, $\xi_m$, $m \in \{1,2,\ldots,M\}$. The SOM weight vectors are updated using the recursive updating relation

$$\xi_m(k+1) = \xi_m(k) + \xi_m(k) h(k, \text{dis}(m, b(k))) [\bar{s}_m - \xi_m(k)]$$

(7)

where $k$ is the index of the training item $s(k)$, $\text{dis}(m, b(k))$ is the topological distance between region $m$ and $b(k)$ computed using the Breadth-first procedure [18] and $\bar{s}_m$ is the sample mean of the training vectors in region $m$. For each training item, the Best Matching Unit (BMU) is index of the local model for which $s(k) \in V_m$, or

$$b(k) = \arg\min_m ||s(k) - \xi_m||$$

(8)

The function $h(k, \text{dis}(m, b(k)))$ is the neighborhood function that enables each vector to be updated using training samples in neighboring regions

$$h(k, \text{dis}(m, b(k))) = \exp\left(-\frac{\text{dis}(m, b(k))^2}{2\sigma^2(k)}\right)$$

(9)

where the width parameter, $\sigma(k)$, is usually taken to be larger in the initial stages of training and tends to zero as $k \to \infty$ to achieve convergence and global ordering of the SOM [19]. In this paper, $\sigma(k)$ was decreased linearly with each pass through the training data (epoch).

Since the ultimate goal of the SOM in the GSMMS is to achieve accurate modeling, the updating of SOM nodes described by Eq. 7 is augmented with a penalty term

$$\xi_m(k) = \frac{e_m(k)}{\sum_{m=1}^{M} e_m(k)}$$

(10)

where $e_m(k)$ are the RMS modeling errors in the $m$th region. By examining Eq. 7, one can see that due to the term $\xi_m(k)$, the weight vectors will tend to move towards regions with higher modeling errors. This ultimately leads to a finer partition of the state space in areas of high nonlinearity.

The local model parameters in region $m$ are determined by minimizing

$$J_m(\theta_m) = \frac{1}{k} \sum_{i=1}^{k} w_m(s(i)) ||y(i) - \hat{y}_m(i)||^2$$

(11)

where $\theta_m$ denotes the model parameters in region $m$, $y(i)$ is the training output at time $i$ and $\hat{y}_m(i)$ is the predicted output of regional model $m$ at time $i$. The function $w_m(s(i))$ determines the weight of the modeling error associated with sample $i$ for the
cost function in region \( m \), allowing each training sample \( s(i) \) to update a neighborhood of regional models. The weighting function is defined as

\[
w_m(s(i)) = \exp \left( -\frac{\text{dis}(m, b(k))^2}{2\sigma(k)^2} \right)
\]  

(11)

so that training samples have a larger effect on the model parameters near the BMU and a smaller effect farther away from the BMU. This cooperation allows the GSMMS to utilize each training sample to affect all local models. Such regional cooperation has been shown to have the ability to speed convergence in the early stages of training as well as increase the bias in later training states [15] (which is why \( \sigma(k) \) was reduced to narrow \( w_m \) with each subsequent epoch).

After a predetermined number of passes through the training data (epochs), the local model with the highest modeling error is selected. The state space partition is then refined by adding a node near the poorly modeled region. The growth mechanism essentially follows the growing cell structure method proposed in [13] and [14], but uses modeling errors instead of visitation frequency or quantization errors as an insertion criteria.

Finally, a stopping criteria was applied to terminate the training. In this paper, two stopping criteria were used. If the total RMS error was below a pre-determined tolerance, or if the number of SOM nodes exceeded a pre-determined number, the training was terminated.

### 2.3 Interpretation of Residuals in GSMMS Based Anomaly Detection

Anomaly detection is accomplished through comparison of the statistical characteristics of the GSMMS modeling residuals displayed during normal behavior with the characteristics of the most recent residuals. The modeling residuals are defined as differences between the system output and the output of the GSMMS describing the normal system behavior. Clearly, one would expect changes in the system dynamics to result in changes in the behavior of the modeling residuals. However, interpretation of the residuals is not a trivial task. A GSMMS model, like any other “divide and conquer” model, will have regions with different levels of approximation accuracy. Thus, surges in the modeling residuals could occur due to changes in the operating region alone, resulting in false alarms. Therefore, it is necessary to have an anomaly detection strategy that accounts for the regionally dependent residual structures.

An important advantage of the multiple model system framework is that the state space is decomposed into regions of similar dynamic behavior. Just as the GSMMS decomposes the modeling into simpler local models, the residual interpretation can be conducted on the simpler regional residuals. Thus, each region of the GSMMS is equipped with its own decision making scheme that quantifies how close the current residual pattern is to the normal pattern.

Following [3], the performance within each region \( m \) was quantitatively described using the concept of regional confidence values (CVs) defined as

\[
CV(m,k) = \frac{|f_m(e) \cdot g_m(e,k)|}{\|f_m(e)\| \|g_m(e,k)\|}
\]  

(12)

where \( f_m(e) \) is the probability density function (PDF) of the modeling residuals displayed during normal behavior and \( g_m(e,k) \) is the PDF of the residuals corresponding to the current behavior at time \( k \). The regional confidence value, \( CV(m,k) \), describes a normalized area of overlap of the PDFs in that region. It is easy to see from Eq. 12 that \( CV(m,k) = 1 \) if the current residual PDF is exactly the corresponding normal behavior PDF and less than 1 otherwise. The PDF \( f_m(e) \) was approximated using Gaussian Mixture Models due to their universal approximation capability [20] and \( g_m(e,k) \) was calculated by updating \( f_m(e) \) recursively during operation [21]. Using the Gaussian Mixture Model formulation, a closed form solution for (12) can be obtained to quickly compute \( CV(m,k) \) at each time step, \( k \).

To simplify the monitoring scheme, the regional CVs were merged into a global CV defined as the geometric mean to emphasize individual departures from normal behavior (CV = 1) since a decrease in any CV indicates that the system is behaving abnormally (at least, in the region where the local CVs are low).

### 2.4 Anomaly Isolation Through Distributed Anomaly Detection

Isolation of the anomaly source can be conducted by reconfiguring and reconnecting anomaly detectors (ADs) to subsystems in the anomalous system. This approach is depicted in Fig. 1 for a generic system. The overall AD of the anomalous system is replaced with ADs that monitor the constituent subsystems. Only the AD of the anomalous subsystem will be low. Clearly, for complex systems of interacting subsystems, this process could be repeated until anomalous subsystems are isolated to the smallest possible granularity as illustrated in Fig. 1.

![Illustration of Fault Isolation Through Distributed Anomaly Detection](image-url)
Once the anomaly is isolated, the next step is to determine whether the detected anomaly has been observed before based on a set of corresponding training signatures, or if the fault is an unknown fault that has never been encountered. As stated in Section 2.3, anomaly detection is accomplished by comparing current system behavior to a GSMMS of normal behavior and computing CVs defined by Eq. 12. The same approach can be applied to construct a diagnoser for a specific fault. Data emitted in the presence of a certain type of fault can be used to train a GSMMS model and the region dependent modeling residual PDFs in the presence of the fault can be estimated.

Presence of that fault can be detected when the most recent modeling residuals generated by the corresponding GSMMS match the modeling residuals observed during training of that GSMMS. Following [22], the similarity of modeling residuals will be evaluated in terms of the appropriate diagnoser CV. This process is similar to how the CV of the normal behavior is evaluated based on the overlaps between the most recent modeling residuals and those generated by the normal behavior GSMMS. Essentially, in the presence of a previously known and modeled fault, the appropriate diagnoser will display high CV levels (its modeling residuals match well with those observed in the presence of that fault), while all others will display low CVs.

If none of the diagnosers are able to identify its corresponding fault with a high degree of confidence (all CVs are low), the fault is unknown and can be diagnosed in the future by training a new GSMMS for the currently observed system behavior.

3 APPLICATION TO THE EGR VALVE SYSTEM

Exhaust Gas Recirculation (EGR) is widely used as a method to reduce NO\textsubscript{x} emissions. In an EGR system, a portion of the exhaust gas is introduced into the intake. This proportion is controlled by the EGR valve which is set by the engine control unit based on the current operating conditions (engine speed and load). A simplified schematic of the EGR system is shown in Fig. 2 and the block diagram of the EGR valve system is shown in Fig. 3. The GSMMS based anomaly detection, isolation and fault diagnosis procedure was applied to the EGR valve system of a four cylinder turbo charged diesel engine. The software package En-DYNA THEMOS CRTD 2.0 by TESIS was used to conduct the engine simulations and the throttle was modeled as an isentropic flow [23]. The controller consists of two look up tables determining the mass flow \(m_{\text{desired}}\) and the coarse control for the throttle angle. A proportional-integral (PI) controller is then used to correct the throttle angle such that the desired mass flow is achieved [23]. The look-up table values and PI controller parameters were provided with the default calibration.

![Figure 3. BLOCK DIAGRAM OF THE EGR VALVE SYSTEM](image)

3.1 Modeling Procedure

En-DYNA was used to generate engine data and the signals were polluted with 2% additive noise. For each of the blocks in Fig. 3, a GSMMS for normal behavior was constructed. Three standard driving profiles (ECE-15, FTP-75 and MVEG-B) were used to generate training data and a fourth profile (Japan 10-15) was used to test the generalization capability of the model.

The GSMMS models of normal and faulty behavior modes were trained offline using inputs and outputs from relevant subsystems. For each GSMMS, the appropriate inputs were selected and autoregressive orders of the local models were provided. One should note that recent work has shown that even input selection can be automated [24], further decreasing the requisite a priori knowledge. The training procedure followed the batch algorithm from [2], with the stopping criteria being a maximum SOM size. However, a few additional modifications were used for this application. It was found during the modeling procedure that inserting nodes based solely on root-mean square error as was done in [2] resulted in regions in which the errors were not normally distributed. This indicated a possibly poor approximation since local model parameters were estimated using a least squares approach. To improve modeling accuracy and push modeling errors to follow a Gaussian distribution, a new node was inserted between the region \(m\) satisfying

\[
\max_m \text{card}\{e_m : |e_m| > 4\sigma_m\}
\]

and its furthest neighbor (in Euclidean distance). In the above equation, card \{\cdot\} denotes the set cardinality, \(e_m\) are the modeling errors and \(\sigma_m\) is the standard deviation of the errors for region \(m\). Since the \(\pm 4\sigma\) limits contain over 99.99% of the data for a normal distribution, the above criteria can be considered a measure of local residual non-normality.

\[\text{Figure 3. BLOCK DIAGRAM OF THE EGR VALVE SYSTEM}\]

\[\text{Figure 2. SCHEMATIC OF A GENERIC EGR SYSTEM}\]
3.2 Anomaly Detection and Isolation

Several abnormalities were introduced into the plant and controller of the EGR valve system. The methodology described in Section 2 was then used to detect and isolate the anomalies with no prior knowledge of the faults or training based on the anomalous behavior data.

An anomaly was introduced to the EGR valve by modifying the throttle characteristic curve that describes the mass airflow across the throttle as a function of the throttle angle, $\alpha$. Fig. 4 shows the faults that were introduced as anomalies to the EGR valve. Fig. 5 shows the AD that is responsible for anomaly detection and Fig. 6 shows the configuration of the ADs that were used for anomaly isolation. The first anomaly detector (AD1) was set to monitor the entire EGR valve system. AD2 and AD3 monitored the look-up tables. AD4 monitored the PI controller and AD5 was set to track the EGR valve behavior. The results of AD2 through AD5 are used to isolate the fault to the subsystem with the smallest possible granularity, once AD1 detected an anomaly.

The global CVs for AD1 can be seen in Fig. 7 while the global CVs for AD2 through AD5 can be seen in Fig. 8. It is obvious from these figures that interpretation of AD1 through AD5 localizes the fault to the EGR valve, even though no prior training of models for faulty behavior was done.

An interesting observation in Fig. 8 is that AD1 is not as sensitive to plant faults as the valve anomaly detector, AD5. The reason for this is that AD1 monitors the controlled EGR valve system and the PI controller compensates for some of the adverse effects of the valve anomaly.

A controller anomaly was simulated by introducing a delay into the control loop as indicated in Fig. 10. The increased delay emulates a situation when the control system communication is degraded. Once again, only AD1 is monitored until an anomaly is detected and AD2 through AD5 are utilized to isolate the fault. Fig. 9 shows the CV for AD1 and clearly indicates an anomaly.
The above anomaly detection and isolation procedure follows the methodology described in Section 2 with one notable modification. Since the EGR valve angle saturates \(0^\circ \leq \alpha \leq 90^\circ\), the PI controller output has a saturation nonlinearity. Therefore, when the EGR valve is faulty (restricted), the PI controller will attempt to correct to achieve the desired \(\dot{m}_{EGR}\). Thus, regions of the normal PI GSMMS that did not achieve saturation during training will now reach saturation, which will change the PDF of the residuals of the GSMMS modeling the PI controller and result in a drop in the CV for \(\text{AD}_4\). To avoid this, the CV for the PI controller was not updated when the EGR valve was operating near the saturation limits. This had the effect of limiting the number of regions that would reach saturation during anomalous plant behavior. While this strategy proved to be effective, the updating “threshold” (how far from saturation \(\alpha\) needed to be to proceed with updating the CVs of the PI controller) was determined mostly by trial and error.

3.3 Fault Diagnosis

The same GSMMS model-based framework used for anomaly detection and isolation is used to diagnose (recognize) faults using the methodology described in Section 2.4. Data emitted by faulty systems was used to construct 6 diagnosers (Ds). \(D_1\) through \(D_3\) were trained to diagnose faults on the EGR valve using data from valve faults 1 through 3 respectively (See Fig. 4). \(D_4\) through \(D_6\) were trained to diagnose 10, 15 and 20ms delays of the EGR valve controller. The D configuration is illustrated in Fig. 12. Fig. 13(a) shows the CVs of the 3 Ds relevant to the EGR valve when fault 3 was present while Fig. 13(b) shows the controller Ds when the 20 ms delay was present. As expected \(D_3\) and \(D_6\) have the highest CVs indicating that the associated GSMMS models are the nearest in terms of modeling the particular anomalous system behavior. As mentioned in Section 2.4, if the anomaly is unprecedented all Ds would indicate low CVs and a new GSMMS would have to be trained to model the new anomalous dynamics.

4 CONCLUSIONS AND FUTURE WORK

In this paper, a recently introduced method for precedent-free anomaly isolation based on distributed anomaly detection is applied to an EGR valve system. The “divide and conquer” GSMMS modeling approach was successfully employed to construct models for the EGR valve system and its constituent subsystems. Based on these models, the presence of anomalous behavior in various subsystems was detected and the sources of the anomalies were isolated without the need for prior training with data characterizing the underlying fault. Finally, the same approach based on characterization of GSMMS modeling residuals was used to recognize previously encountered (and modeled) faults.
Future work will focus on two areas. First, the GSMMS of the faulty behavior and the local analytical tractability of the GSMMS may be exploited to apply a control scheme that can compensate for the anomalous behavior. Additionally, an area of future work will be to implement the methodology of this paper on a real engine.

REFERENCES


