Networked vehicles for automated fault detection

Stefan Byttner, Thorsteinn Rögnvaldsson, Magnus Svensson, George Bitar and Wesley Chominsky

N.B.: When citing this work, cite the original article.

Byttner S, Rognvaldsson T, Svensson M, Bitar G, Chominsky W. Networked vehicles for automated fault detection. In: 2009 IEEE International Symposium on Circuits and Systems: circuits and systems for human centric smart living technologies, conference program, Taipei International Convention Center, Taipei, Taiwan, May 24-May 27, 2009. Piscataway, N.J.: IEEE; 2009. p. 1213-1216.

DOI: http://dx.doi.org/10.1109/ISCAS.2009.5117980
Copyright: IEEE

Networked vehicles for automated fault detection

S. Byttner*, T. Rögnvaldsson*‡, M. Svensson†, G. Bitar§, W. Chominsky¶
*Intelligent Systems Laboratory, Halmstad University, SE-301 18 Halmstad, Sweden
†Volvo Technology, SE-405 08 Göteborg, Sweden
‡AASS, Örebro University, SE-701 82 Örebro, Sweden
§Volvo Technology of America, 7825 National Service Rd., Greensboro, NC 27409, USA
¶Volvo Trucks North America, 7900 National Service Rd., Greensboro, NC 27409, USA

***Abstract*— Creating fault detection software for complex mechatronic systems (e.g. modern vehicles) is costly both in terms of engineer time and hardware resources. With the availability of wireless communication in vehicles, information can be transmitted from vehicles to allow historical or fleet comparisons. New networked applications can be created that, e.g., monitor if the behavior of a certain system in a vehicle deviates compared to the system behavior observed in a fleet. This allows a new approach to fault detection that can help reduce development costs of fault detection software and create vehicle individual service planning. The COSMO (COnsensus Self-organized MOdeling) methodology described in this paper creates a compact representation of the data observed for a subsystem or component in a vehicle. A representation that can be sent to a server in a backoffice and compared to similar representations for other vehicles. The backoffice server can collect representations from a single vehicle over time or from a fleet of vehicles to define a norm of the vehicle condition. The vehicle condition can then be monitored, looking for deviations from the norm.**

The method is demonstrated for measurements made on a real truck driven in varied conditions with ten different generated faults. The proposed method is able to detect all cases without prior information on what a fault looks like or which signals to use.

I. Introduction

Two things that are very important for maximizing vehicle uptime while minimizing cost are that the vehicle is operated correctly and that it is maintained properly with respect to how it is being used. The vehicle operation and the maintenance schedule are intimately connected; a change in operational profile often results in a change in maintenance schedule to maintain optimal utilization. Operation and performance characteristics should therefore be monitored frequently to allow optimal use of the vehicle. However, this is very costly and only the most expensive vehicles (e.g. airplanes) can be equipped with special sensors and software to monitor the individual equipment health and usage during operation in order to maximize uptime and minimize maintenance cost. It is, however, no less important to monitor the individual health of more common and less costly vehicles, like city buses and trucks. A considerable fraction of the worldwide downtime logistics costs are probably due to faults that occur on vehicles that are more common than very expensive transportation vehicles (there are probably more than ten million commercial trucks in the world and about five million commercial buses, whereas there are probably only a few hundred thousand commercial aircraft).

However, building a feasible health monitoring system for a bus or a truck, which are produced in large quantities and extra sensors to monitor vehicle status are often out of the question for cost reasons, requires an approach that is different from the current standard in equipment health monitoring. The present state-of-the-art paradigm in on-line performance monitoring, which has been developed mostly for monitoring large process plants and airplanes, of equipment and systems is to model the process or signal values as a function of other signals and then compare the signal values predicted by the model with measured values during operation [1]. The modelling is done off-line, often involving a human expert modeller that does model selection and validation, and can be done with a variety of methods, e.g. kernel regression, non-linear partial least squares, neural networks, or any other preferred nonlinear method.

The approach we take in this paper is different. Instead of comparing a measured signal at one time to a model built on data collected at a different time, we propose to compare models built at these different times. If the models disagree, then this is an indication that something has changed. Also, instead of doing the modelling off-line we propose a method where the modelling and model selection is done automatically and on-board the vehicle during operation. The approach can be described as producing compressed images of the data observed on the vehicles over a certain time window and compressed images from the same vehicle at different times (or from different vehicles) are compared. If the image has changed (or is different from the other vehicles images) then this is an indication of a possible fault.

The main reason for using models in the comparison is to be able to transmit the data over a limited bandwidth telematics gateway to a back office application that can do data mining on the observed data. The reason for having an automated modelling method is because a self-organizing system that can automatically handle several vehicles in a fleet and detect previously unknown errors without being explicitly told about them is necessary, or else the chance for commercial implementation in ordinary commercial vehicles is low.

The approach we use here to compress the data is to build linear models for the relationships between the signals on the vehicle. Linear models were chosen for their simplicity,

978-1-4244-3828-0/09/$25.00

the absence of local minima, the fact that they can easily be implemented in hardware, and the empirical observation that relationships between sensor signals in condition monitoring situations often end up being almost linear [1]. The interestingness of a relationship is measured by the error, the sum of the squared residuals, for the model. A low error means that there is a strong (linear) relationship, which we interpret as a possibly interesting relationship that we want to monitor over time. The parameters of interesting models are sent via a telematics gateway to a back office application that compares the models from different vehicles (or models for the same vehicle over time). The idea is that strong correlations between sensor values hold important information about the function of the system. This idea is supported by a recent paper [2] where a diagnostics method is proposed based on correlations and where it is showed that they contain relevant information for fault diagnosis. However, in [2] are the observed correlations given beforehand, there is no automatic search for good correlations, and the correlations are compared to expected values determined beforehand in a laboratory, not against correlations measured on a fleet of similar vehicles under similar conditions.

We have previously presented and demonstrated our method on simulated data for detecting leakage in a simulated cooling system [3] and faults in an air suspension system [4]. This paper presents the results of applying the method to more extensive measurements on a Volvo truck with ten different fault conditions related to the engine system.

II. Description of measurements

The measurements were made on a Volvo VN780 with a D12D engine. The sensor data collected on the J1587 vehicle data bus included:

- Turbo Boost Pressure
- Exhaust Gas Recirculation (EGR) Valve 1 Position
- Exhaust Gas Recirculation (EGR) Gas Temperature
- Accelerator Pedal Position
- Actual Engine Percent Load
- Engine Coolant Temperature
- Instantaneous Fuel Rate

And two additional sensors:

- Charge Air Cooler (CAC) Inlet Temperature
- Charge Air Cooler (CAC) Inlet Pressure

Data was collected for a four hour test route with a wide variety of driving conditions. All signals were measured with a one second sampling interval. Four different runs were measured with normal driving conditions (i.e. no fault was introduced), and ten different runs with specific faults. The faults are leaks and restrictions introduced at the air intake, turbo inlet, CAC inlet, and exhaust. During the fault data collection period, fault levels were determined that ranged from those just noticeable by the driver, to levels undetectable by the driver. The different faults were:

- Charge Air Cooler leaks with hole diameters of 3/32”, 3/16” and 3/8”
- 5” exhaust pipe downsized - diameters of 4”, 3.5” and 3”
- Air filter blocked by 33% and 50%
- Front grill blocked by 50% and 100%

These faults were chosen because they represent significant reasons for unplanned downtime in the field, and they could be introduced without extensive modification.

III. COSMO fault detection

Our approach, COnsensus Self-organized MOdels (COSMO) [3][4] is based on creating model combinations of all signals in a data set, then selecting the models with best fit to be monitored for deviations. For practical reasons, we use linear models and some limitations needs to be made to bound the search space, e.g. limit the number of signals that are allowed in a model and the use of time-lags. In the test we allowed up to three input signals in a model, and no time-lags. This means, with nine available signals, that we have

$$9 \times \binom{8}{1} + 9 \times \binom{8}{2} + 9 \times \binom{8}{3} = 828 \qquad (1)$$

models. All these possible model combinations are tested (this can be done in parallel on a fleet of vehicles) and the models with the lowest *NMSE* are selected, where

$$\mathit{NMSE} = \frac{1}{N\sigma^2} \sum_{n=1}^{N} [y(n) - \hat{y}(n)]^2. \qquad (2)$$

The models with the lowest *NMSE* are then continually monitored in a vehicle (or in multiple vehicles at the same time), and the model parameters can e.g. with certain time intervals be sent to a backoffice server. For each model a comparison is done by doing a leave-one-out test, i.e. with a group of models (the group is formed either by having multiple vehicles or different historic parts of measurements in one vehicle) one model is left out, a probability density for the parameters of the remaining models (denoted reference models) is estimated, and this density is then used to compute a p-value for the left out model's parameters. The procedure is repeated for all the models to get an estimate for each model. The null hypothesis is that the set of model parameters $\mathbf{w}$ follow a multivariate Gaussian distribution, i.e.

$$p(\mathbf{w}) = N(\Sigma) \exp\left[-\frac{(\mathbf{w}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{w}-\boldsymbol{\mu})}{2} \right] \qquad (3)$$

where $N(\Sigma)$ is a normalization factor, $\boldsymbol{\mu}$ is the mean parameter vector, and $\boldsymbol{\Sigma}$ is the covariance matrix for the parameter vectors. When one system is tested, then the remaining systems are used to compute $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ and the p-value for the left out system is computed according to

$$\text{p-value}(\mathbf{w}_{lo}) = 1 - \text{erf}(a) \qquad (4)$$

where erf denotes the error function and

$$a = \sqrt{\frac{(\mathbf{w}_{lo}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{w}_{lo}-\boldsymbol{\mu})}{2}}. \qquad (5)$$

If the p-value is smaller than a certain threshold, then the system with parameters $\mathbf{w}_{lo}$ is flagged as a deviating system (potentially faulty).

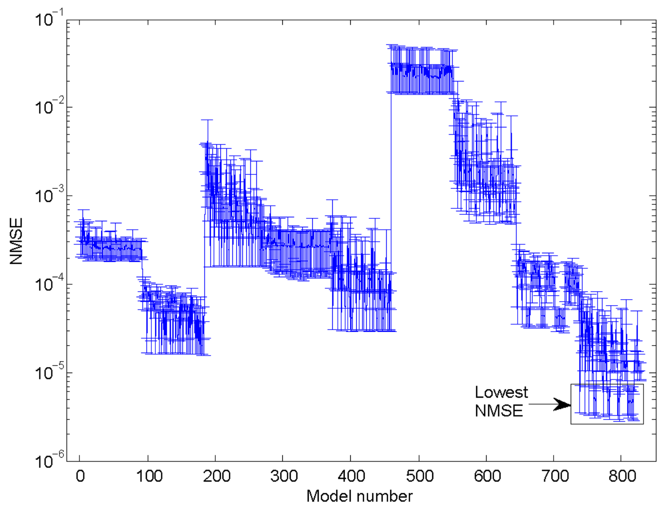

Fig. 1. The normalized mean squared error (NMSE) of the 828 different model combinations with 95% confidence intervals. There is a total of 29 combinations that have the lowest NMSE.

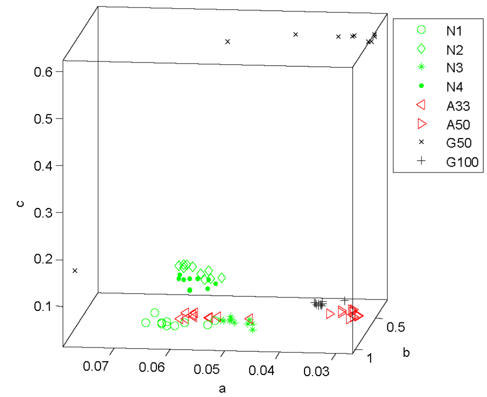

Fig. 2. Model parameters of the model shown in Eq. 6 with the lowest NMSE. N1-N4 are the parameters obtained for normal driving conditons, A33-A50 for the case of blocked air filter and G50-G100 for the case of blocked grill.

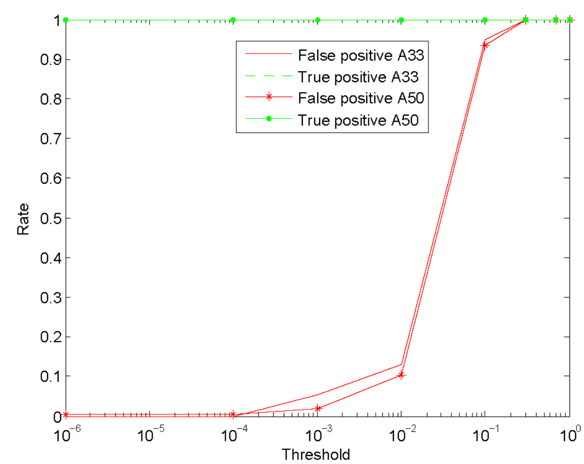

Fig. 3. Threshold value and the rate of classifying a vehicle as deviating for the air filter faults. With a p-value threshold value of 10^{-4} or smaller, there are few false positives and both cases can be detected.

IV. Results

The data set consists of 9 different signals and all signal combinations with up to 3 signals in a model are included, therefore there is a total of 828 different combinations. In Figure 1 the NMSE values of the models and their corresponding 95% confidence interval is shown when they are fitted to data corresponding to normal operation. Of the 828 models, there are 29 models that have significantly lower NMSE than the others (see Fig. 1); out of these models there is one model with one input, seven of them have two signal inputs, and 21 have 3 signal inputs. An example of a model with 3 signal inputs is:

$$P_T = a \cdot (T_{CAC}) + b \cdot (P_{CAC}) + c \cdot (A_P) + d \qquad (6)$$

where P_T is the turbo pressure, P_{CAC} is the CAC pressure, T_{CAC} is the CAC temperature and A_P is the accelerator pedal position. The model parameters $w = [a; b; c]$ (bias excluded) are shown in Figure 2. The figure shows the model parameters for models that have been fitted to the four different datasets of normal operation and four of the fault cases (the other fault cases were excluded to increase readability of the figure). The cases corresponding to normal operation have model parameters which are relatively tightly grouped together. When the faults are introduced the model parameters change (although in the case of air filter 33 the change is not so large).

Ten model instances are created for each of the measurement sets using cross-validation (10% of the data is left out and the 10% that is left out is shifted in a round-robin fashion, thus creating 10 models). For the detection test a group is defined to contain 31 models; The majority of models are assumed to be of normal operating conditions so 30 models are sampled randomly out of 40 possible (there are 40 models of normal operation in total since there are 4 normal measurements with 10 model instances created for each one). One (out of 10) models that were fitted to fault measurements is then added to the group thus creating a group of 31 model instances. This is repeated 10 times since there are 10 model instances for each fault; this results in 10 groups per fault condition, each group containing 31 model instances.

It is then possible to compute the p-values using the leave-one-out test for a group described in the previous section. Figures 3, 4, 5 and 6 show the classification rate for different threshold values on the p-value (for the model in Eq. 6) using a committe of all the 29 models. True positive is when a model is classified as deviating and has been fitted to a faulty data set. False positive is when a model is classified as deviating but has been fitted to a normal (non-faulty) data set. Setting the threshold to smaller values will yield less false positives, but gives a lower detection rate (especially for the case of CAC316). The output of the committe is that a system is deviating if 3 or more models show deviating parameters. Figure 7 shows the rate at which the committe identify the faulty data from the normal data for a specific threshold (p-value $< 10^{-4}$).

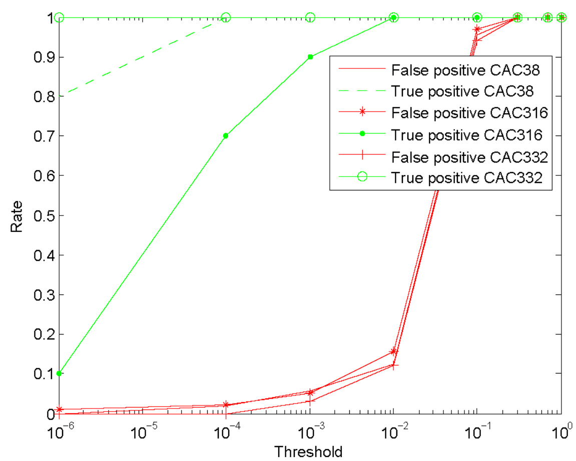

Fig. 4. Threshold value and the rate of classifying a vehicle as deviating for the CAC faults. With a p-value threshold value of 10^{-4} or smaller, there are few false positives and CAC38 and CAC332 can be detected, however CAC316 was harder to distinguish from normal operation (this may be due to mislabeling).

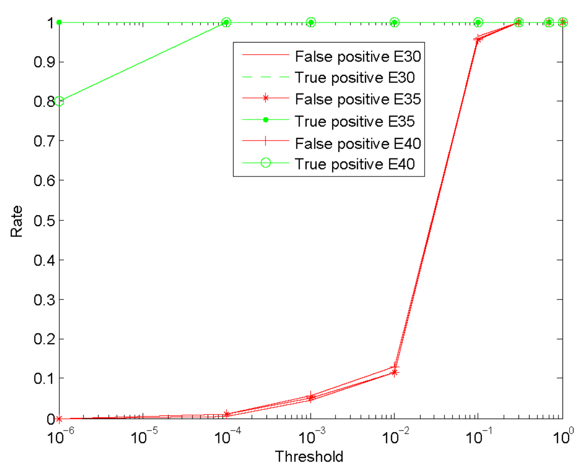

Fig. 5. Threshold value and the rate of classifying a vehicle as deviating for the exhaust pipe downsizing faults. With a p-value threshold value of 10^{-4} or smaller, there are few false positives and all cases can be detected.

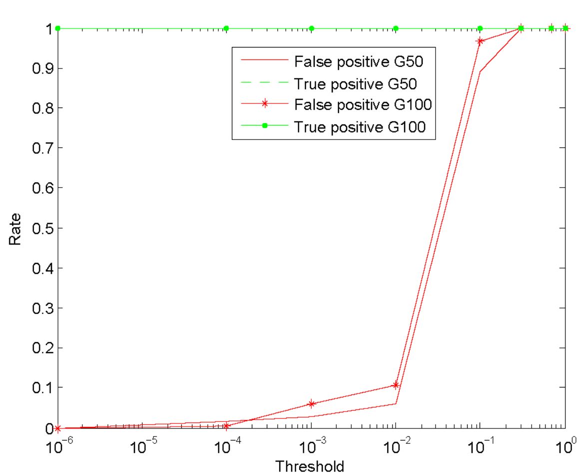

Fig. 6. Threshold value and the rate of classifying a vehicle as deviating for the front grill block. With a p-value threshold value of 10^{-4} or smaller, there are few false positives and both cases can be detected.

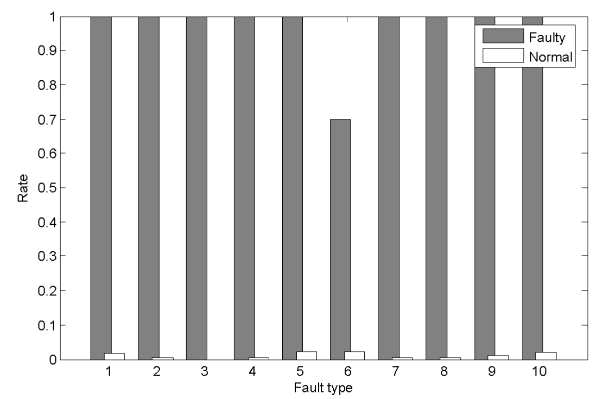

Fig. 7. Rate of classifying as deviating for the two groups; faulty (1) and normal (2) vehicle. The rate is calculated as the median output category of the 29 models. The p-value threshold was chosen as 10^{-4} for all models. Faults 1-2 are grill, 3-4 air filter, 5-7 CAC and 8-10 exhaust pipe faults.

V. Conclusion

With an increased focus on managing up-time and improving vehicle service planning there is a need to get vehicle individual health information. As complexity in vehicles increase (amount of electronics and functionality) it becomes more difficult and costly to create fault detection software. At the same time there are large amounts of data available on the data links inside vehicles, but there are not many tools to extract useful information from them. A self-organized methodology (COSMO) has been developed that searches for linear models (correlations) between signals, and monitors the model parameters for deviations. The strategy can take advantage of wireless communication to do real time data mining on vehicles while they are being used, and there is no need to pre-specify the characteristic of the fault. The paper shows the results of applying the method to data collected during normal operation of a vehicle, and for ten different fault cases. The method managed to identify all fault cases as deviating while only making few false positives (classifying a non-faulty vehicle as deviating). To be able to discriminate between the different faults, it would e.g. be possible to build a knowledge base of faults based on feedback from a mechanics workshop. This knowledge base could map parameter deviations to specific faults, allowing for tracking of deterioration of components or subsystems in vehicles.

References

[1] J. W. Hines and R. Seibert, "Technical review of on-line monitoring techniques for performance assessment. volume 1: State-of-the-art," U.S. Nuclear Regulatory Commission Office of Nuclear Regulatory Research, Washington, DC 20555-0001, Report NUREG/CR-6895, 2005.

[2] S. H. D'Silva, "Diagnostics based on the statistical correlation of sensors," Society of Automotive Engineers (SAE), Technical paper 2008-01-0129, 2008.

[3] S. Byttner, T. Rögnvaldsson, and M. Svensson, "Self-organized modeling for vehicle fleet based fault detection," Society of Automotive Engineers (SAE), Technical paper 2008-01-1297, 2008.

[4] ——, "Modeling for vehicle fleet remote diagnostics," Society of Automotive Engineers (SAE), Technical paper 2007-01-4154, 2007.