A combined diagnosis system design using model-based and data-driven methods

Daniel Jung, Kok Yew Ng, Erik Frisk, and Mattias Krysander

***Abstract*— A hybrid diagnosis system design is proposed that combines model-based and data-driven diagnosis methods for fault isolation. A set of residuals are used to detect if there is a fault in the system and a consistency-based fault isolation algorithm is used to compute all diagnosis candidates that can explain the triggered residuals. To improve fault isolation, diagnosis candidates are ranked by evaluating the residuals using a set of one-class support vector machines trained using data from different faults. The proposed diagnosis system design is evaluated using simulations of a model describing the air-flow in an internal combustion engine.**

I. Introduction

Two common approaches in fault diagnosis are usually referred to as model-based diagnosis [14] and data-driven diagnosis [15]. Both approaches use models to detect and isolate faults that occur in the system. In general, model-based diagnosis uses physical models of the system and modelled faults, while data-driven diagnosis uses models learned from training data.

A general diagnosis system design in model-based diagnosis is to use a set of residuals, computed based on different parts of the model, to detect if a fault has occurred in the system [2]. A residual is said to have triggered if it deviates from its nominal behavior. Based on the triggered residuals, all diagnosis candidates that can explain the triggered residuals using a consistency-based fault isolation algorithm are computed [3].

In many applications, false alarms are troublesome and will result in the true diagnosis candidate being rejected. Therefore, thresholds are selected such that the false alarm rate is very small. As such, not all residuals will trigger as expected if the fault is small. Thus, even though all faults are uniquely isolable given the set of residuals in the ideal case, it is not certain all residuals will trigger when a fault occurs. As an example, consider the fault signature matrix in Table I where an X at position (i, j) means that the residual r_i is sensitive to the fault f_j. If only r_1 and r_2 trigger, the possible conclusions that can be made are either f_1 or f_2 has occurred.

Having said that, even though both f_1 and f_2 can explain the two triggered residuals, the residual values for each fault

TABLE I
Example fault signature matrix.

	f_1	f_2	f_3	$\dots$	f_{n_f}
r_1	X	X	X	$\dots$	0
r_2	X	X	0	$\dots$	X
$\vdots$					$\vdots$
r_{n_r}	0	X	X	$\dots$	0

might be different. By analyzing the correlation between the residuals it could be possible to improve fault isolability performance and isolate the two faults from each other, or at least tell which of the faults is more likely to have occurred, even before all residuals sensitive to the fault have triggered. As an illustrative example, assume that when analyzing residuals r_1 and r_2, the residual values for each of the two single-fault cases will be located as shown in Fig. 1. The dashed lines represent thresholds and the highlighted region at the center represents the nominal residual behavior, i.e. no residual has triggered. In this example, residual data belongs to the model of fault f_1 which is a more likely diagnosis candidate with respect to f_2.

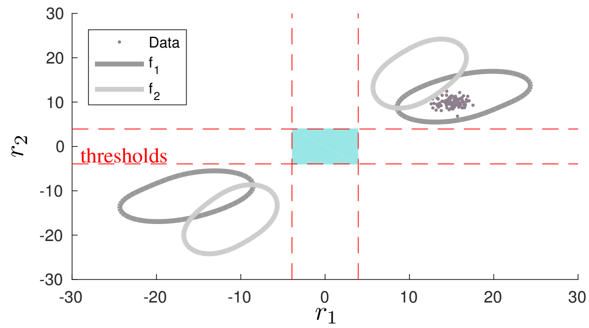

Fig. 1. An example where two different faults will trigger the same set of residuals but different residual outputs.

If there are several diagnosis candidates that can explain the triggered faults, it might not be obvious which possible set of faults is present in the system. This will make it more difficult as well as time-consuming and costly for the control system to take suitable counter-measures or for an technician to troubleshoot the system. It is not desired to reject too many diagnosis candidates if there is a risk that the true fault might also be rejected. By analyzing the correlation between residuals and using knowledge from previous faults, it would be possible to identify which diagnosis candidates are more likely to have occurred. This is relevant, for example, when designing diagnosis systems for automotive

This work was supported by Volvo Car Corporation in Gothenburg, Sweden.

D. Jung, E. Frisk, and M. Krysander are with Department of Electrical Engineering, Linköping University, Linköping, Sweden. (Emails: daniel.jung@liu.se, erik.frisk@liu.se, mattias.krysander@liu.se)

K.Y. Ng is with Electrical and Computer Systems Engineering, School of Engineering, Monash University, 47500 Bandar Sunway, Malaysia. (Email: ng.kok.yew@monash.edu)

applications since the troubleshooting can be performed more efficiently at the workshop.

This paper proposes a combined diagnosis system design to make use of the advantages from both model-based and data-driven fault diagnosis approaches. The purpose is to improve fault isolation in situations when the fault is not yet isolated by ranking the diagnosis candidates given which is more likely. The proposed method can handle both single-faults and multiple-faults, and also identify when it is likely that an unknown fault has occurred.

Previous works that also combined model-based and data-driven fault diagnosis techniques are, for example, [4], [13], and [8]. In [7] and [8], different test quantities are designed using model-based and data-driven methods. Combined model-based and data-driven methods have also been proposed for prognostics and condition-based maintenance [1], [10]. With respect to previous works, a combined fault isolation strategy is proposed which combines the advantages of both consistency-based fault isolation and data-driven fault modeling.

II. Problem statement

The goal is to develop a hybrid diagnosis system design utilizing both model-based and data-driven diagnosis methods to better identify the faults that are present in the system. By including data-driven classifiers trained using available data with known faulty behaviors, it should be possible to rank different diagnosis statements which is more likely based on experience from previous faults.

It is assumed that there is an available set of residuals that has been developed to detect and isolate a set of faults in the system. The threshold for each residual is tuned such that the false alarm rate is negligible. However, it is assumed that all residuals might not always trigger when there is a fault, such as when the fault is small, and as such the exoneration assumption is not valid. All diagnosis candidates that can explain the triggered faults are computed using the consistency-based fault isolation algorithm described in [3].

During the development of the diagnosis system, a candidate set of faults are taken into consideration in the diagnosis system design. However, new types of faults can occur that are not included among the fault candidates and this could result in drawing the wrong conclusions about the system state if these unknown faults are not taken into consideration in the fault isolation process. Thus, an important task in the fault isolation process is to identify when it is likely that an unknown fault has occurred.

III. Modeling fault modes for fault isolation

Consider a system to be monitored and let $\mathcal{F} = \{f_1, f_2, \dots f_{n_f}\}$ be a set of possible faults. To describe the system state, the term *fault mode* is used. A fault mode $F \subseteq \mathcal{F}$ is a set of faults present in the system. The fault mode describing the fault-free system $F = \emptyset$ is the nominal state and is denoted by NF (No Fault). Let $\mathcal{R} = \{r_1, r_2, \dots, r_{n_r}\}$ represent the set of residuals that are computed based on the sensors and actuators data from the system. The residuals are designed such that different residuals are sensitive to different subsets of faults. If a residual r_k is not sensitive to a fault f_i, it is said that f_i is decoupled from r_k. Also, let $R_{F_j} \subseteq \mathcal{R}$ denote the subset of residuals where the faults F_j are decoupled.

One approach to compare different fault isolation approaches is to analyze the assumptions regarding fault modes that are utilized in each approach [6]. To perform the analysis, the set of residual outputs that can be explained by different fault modes are modeled. Let $\Phi_R(F_i) \in \mathbb{R}^{\dim(R)}$ where $R \subseteq \mathcal{R}$ denotes the space of all values of the subset of residuals R that can be explained by a fault mode F_i.

Then, each fault mode F_i can also be represented by the set of possible residual values $\Phi_R(F_i)$ that can be explained by F_i. It is assumed that different sets of $\Phi_R(F_i)$ for different fault modes F_i can be overlapping, i.e. the exoneration assumption is not valid. This is the case, for example, if faults are assumed to be small and difficult to detect. The sets $\Phi_R(F_l)$ for each fault mode are seldom completely known, and different assumptions are made about the different $\Phi_R(F_l)$ depending on how the diagnosis system is designed [6].

A. Consistency-based fault isolation

A common approach in model-based diagnosis to perform fault isolation is to use a set of residuals that are sensitive to different sets of faults. If a fault is decoupled from a residual, then the residual is not affected by the presence of the fault when the fault occurs. Based on which residuals have triggered, diagnosis candidates that can explain the triggered residuals are computed. A diagnosis candidate is called minimal if no subset of faults can explain the triggered residuals [3]. Thus, a diagnosis candidate is rejected if it cannot explain the triggered set of residuals.

In consistency-based diagnosis, a set of faults $F \subseteq \mathcal{F}$ is a diagnosis candidate unless a residual where all faults in F are decoupled has triggered [3]. The fault isolation algorithm in [3] computes all minimal diagnosis candidates, i.e. all diagnosis candidates of which no subset of faults is a diagnosis candidate. For example, if no residual has triggered, then all fault combinations are diagnosis candidates. However, the fault-free case is the minimal diagnosis candidate.

1) Modelling fault modes: The approximation of each $\Phi_{\mathcal{R}}(F_l)$ based on the consistency-based approach is denoted by $\Phi^{cb}_{\mathcal{R}}(F_l)$ and is defined as follows: Let J_i be a threshold such that a residual r_i is said to have triggered if $|r_i| > J_i$. Then, the set $\Phi_{\mathcal{R}}(F_l)$ is approximated as $\Phi_{\mathcal{R}}(F_l) = \mathbb{W}_1 \times \mathbb{W}_2 \times \dots \times \mathbb{W}_i \times \dots \times \mathbb{W}_{n_r}$ where

$$\mathbb{W}_i = \begin{cases} \mathbb{R} & \text{if } r_i \text{ is sensitive to any fault } f_j \in F_l \\ [-J_i, J_i] & \text{otherwise.} \end{cases} \tag{1}$$

Note that if $F_1 \subseteq F_2$, then $\Phi^{cb}_{\mathcal{R}}(F_1) \subseteq \Phi^{cb}_{\mathcal{R}}(F_2)$ [6].

The approximation of the feature set $\Phi_{\mathcal{R}}(F_l)$ in (1) describes a set which has orthogonal boundaries with respect to residuals where all faults in the fault mode are decoupled. Note that the fault-free mode NF is described by an n_r dimensional hypercube bounded in each dimension by the threshold values $[-J_i, J_i]$ as represented by the highlighted area in Fig. 1. This choice of boundary for the NF mode is made because each residual is evaluated individually in the consistency-based fault isolation procedure [3].

Should the thresholds for each residual in $R \subseteq \mathcal{R}$ be selected such that false alarms can be neglected, the approximation $\Phi_R^{cb}(F_l)$ is expected to be conservative with respect to the true $\Phi_R(F_l)$, i.e. $\Phi_R(F_l) \subseteq \Phi_R^{cb}(F_l)$. This is because no assumptions are made about which fault realizations are possible for each fault. If there are no false alarms, no diagnosis candidate is falsely rejected. This is an advantage from the aspect that the correct diagnosis candidate will not be rejected. However, this could result in unnecessary poor fault isolation performance since the number of computed diagnosis candidates might be larger than ideal.

As an example, consider that only residuals r_1 and r_2 in Table I are available and the dashed lines in Fig. 1 represent the thresholds of the two residuals. Then, the fault-free case $\Phi_{r_1,r_2}^{cb}(NF)$ is illustrated by the colored area within the thresholds since the fault-free case is rejected if any residual triggers. The single-fault f_3 only affects r_1, hence $\Phi_{r_1,r_2}^{cb}(f_3)$ is the area within the two horizontal thresholds since f_3 cannot explain if r_2 triggers. For each single-faults f_1 and f_2, i.e. $\Phi_{r_1,r_2}^{cb}(f_1)$ and $\Phi_{r_1,r_2}^{cb}(f_2)$ cover the same set including the whole area since both residuals will be affected by the two faults.

B. Data-driven modeling of fault modes

The consistency-based fault isolation approach is conservative and will produce many possible diagnosis candidates. This will be problematic if the different diagnosis candidates are equally likely to occur since it will be difficult to decide on a suitable counter-measure if it is not clear which faults are present in the system. In order to find a prioritized subset of diagnosis candidates, a data-driven approach is used to classify if residual outputs resemble data from previous faults. A set of training data from previous fault scenarios are used to generate models of the different fault modes.

One candidate data-driven method to estimate each residual set $\Phi_R(F_l)$ where the estimation is denoted by $\Phi_R^{\text{dd}}(F_l)$, is one-class Support Vector Machines (1-SVM) [11]. A classifier $C_l(\bar{r}_i)$ is trained given a set of training data including only data from one specific fault mode F_l to classify if a new sample $\bar{r}_i$ belongs to that fault mode or not. The areas in Fig. 1 given f_1 and f_2 illustrate how $\Phi_R^{\text{dd}}(f_i)$ can look like.

In [11], the 1-SVM is proposed to estimate the support of an underlying probability distribution that the data is drawn from. Then, the boundary of the support is selected such that the probability of a sample taken from the true distribution should be drawn outside of the boundary equals some a-priori false positive rate. This description is suitable for this application where the goal of using 1-SVM is to estimate $\Phi_R(F_l)$ for each F_l using training data from the given fault mode F_l.

If the training data is correctly labeled with respect to the fault mode, one 1-SVM classifier is used to model the training data corresponding to each fault mode included in the training data. Note that one sample of data $\bar{r}_i$ can be a member of different fault modes at the same time. Thus, it is expected that $\Phi_R^{\text{dd}}(F_l)$ will be an approximately optimistic estimation of $\Phi_R(F_l)$ since it is unlikely that training data represents all possible fault realizations given fault mode F_l.

1) One-class Support Vector Machines: A summary of the 1-SVM classifier is presented here. Let $(\bar{r}_1, \bar{r}_2, \ldots, \bar{r}_q)$ be a training set including data from mode F_l where $\bar{r}_i$ is the ith sample of a given subset of residuals $R \subseteq \mathcal{R}$. As described in [12], to train a 1-SVM model, the following optimization problem is solved

$$\begin{aligned} \min_{w,\xi,\rho} \quad & \frac{1}{2}\|w\|^2 + \frac{1}{\nu l}\sum_i \xi_i - \rho \\ \text{s. t.} \quad & w \cdot \Psi(\bar{r}_i) \geq \rho - \xi_i, \xi_i \geq 0, i = 1, 2, \ldots, q \end{aligned} \tag{2}$$

where ν is a tuning parameter related to the smoothness of the boundary of the data set. Then, the 1-SVM classifier for a new sample $\bar{r}$ is given by

$$C_R^{F_l}(\bar{r}) = \begin{cases} 1 & \text{if } \sum_i \alpha_i k(\bar{r}_i, \bar{r}) > \rho \\ 0 & \text{otherwise,} \end{cases} \tag{3}$$

i.e., $C_R^{F_l}(\bar{r}) = 1$ if $\bar{r}$ belongs to the class and zero otherwise, where $k(\bar{r}_i, \bar{r})$ is the Gaussian kernel

$$k(x, y) = \Psi(x) \cdot \Psi(y) = e^{-\|x-y\|^2/c} \tag{4}$$

and the parameters α_i are given by the following dual problem

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j k(x_i, x_j) \\ \text{s. t.} \quad & 0 \leq \alpha_i \leq \frac{1}{\nu l}, \sum_i \alpha_i = 1. \end{aligned} \tag{5}$$

IV. METHODOLOGY

Based on the discussion about the consistency-based and data-driven fault isolation approaches, it is expected that the set of computed diagnosis candidates from the consistency-based fault isolation is too conservative as there will be many diagnosis candidates that can make it difficult to identify the true fault. Conversely, a data-driven approach is limited by the quality of the training data and there is a risk where the true diagnosis candidate is rejected. Thus, based on the discussion in Section III, the relation between the true $\Phi_R(F_l)$ and the two approximations: $\Phi_R^{\text{dd}}(F_l)$ and $\Phi_R^{\text{cb}}(F_l)$, can be written as

$$\Phi_R^{\text{dd}}(F_l) \subsetneq \Phi_R(F_l) \subseteq \Phi_R^{\text{cb}}(F_l) \tag{6}$$

for each fault mode F_l. However, if the data-driven models are updated over time as new fault realizations are observed, the lower bound $\Phi_R^{\text{dd}}(F_l)$ should approach the true $\Phi_R(F_l)$.

A fault isolation methodology is proposed to take advantage of both consistency-based and data-driven methods to improve the fault isolation procedure.

A. Combining model-based and data-driven fault isolation

The consistency-based fault isolation algorithm computes all minimal diagnosis candidates, which also represents all possible diagnosis candidates. Each fault mode F_l is modeled as a 1-SVM classifier $C_R^{F_l}$ where $R \subseteq \mathcal{R}$ is the set of residuals used by the classifier. Then, the minimal diagnosis candidates are ranked based on how many of the residual samples when a fault is detected are classified by each corresponding $C_R^{F_l}$. Let $\bar{r}_1, \bar{r}_2, \ldots, \bar{r}_N$, be N samples of the residuals when a fault is detected. If F_i is a minimal diagnosis candidate, its rank is computed as

$$\text{Rank}(F_i) = \frac{\sum_{k=1}^{N} C_R^{F_i}(\bar{r}_k)}{N}, \tag{7}$$

i.e. the percentage of the samples that belongs to F_i. A higher $\text{Rank}(F_i)$ means that the minimal diagnosis candidate F_i is ranked higher.

The 1-SVM classifiers can be interpreted as evaluating new residual outputs using experience from previous faults. Some diagnosis candidates might be prioritized, if the residual data resembles previous observations of the fault mode, but no diagnosis candidates should be falsely rejected. Note that even though faulty data is limited initially when training the 1-SVM classifiers, the performance can be improved as new data is collected over time. This is important, for example, at a workshop where a priority list can be downloaded from the diagnosis system on where to start the troubleshooting from. As data from new faults are collected, the 1-SVM classifiers can be improved over time.

It can also be useful to identify the likelihood that an unknown fault has occurred whereby if the residual output has not been observed before, then it does not belong to any existing fault mode. Let D denote the set of minimal diagnosis candidates, *excluding the unknown fault case*. Then, ranking of an unknown fault F_x is performed as

$$\text{Rank}(F_x) = \frac{\sum_{k=1}^{N} 1 - \bigvee_{\forall F_l \in D} C_R^{F_l}(\bar{r}_k)}{N}, \tag{8}$$

i.e., the percentage of the samples that do not belong to any known fault mode.

B. Modeling multiple-fault modes using single-fault data

Should 1-SMV classifiers be trained for all fault modes representing combinations of multiple-faults, the amount of training data required to train all classifiers grows exponentially. In practice, it is impractical to generate data from all possible multiple-faults scenarios. However, by considering subsets of residuals where the same set of faults are decoupled and then to generate 1-SVM models using these sets, it is possible to evaluate if residual data belongs to multiple-fault modes by only using training data from single-faults.

The key is that a residual with a decoupled fault will not change from its nominal behavior when the fault occurs. Thus, to evaluate if residual data belongs to a double-fault mode $\{f_i, f_j\}$, it is possible to evaluate if residual data of the subset R_j belongs to fault mode f_i and vice-versa, whereby the subset R_j corresponds to residuals where fault f_j is decoupled. If both these cases are true and that each subset of R_i and R_j is non-empty, then the residual output belongs to the fault mode $\{f_i, f_j\}$. Then, the rank of the diagnosis candidate is computed as

$$\text{Rank}(\{f_i, f_j\}) = \frac{\sum_{k=1}^{N} C_{R_j}^{f_i}(\bar{r}_k) C_{R_i}^{f_j}(\bar{r}_k)}{N}. \tag{9}$$

The same approach can be generalized to multiple-faults of higher cardinality.

Fig. 2 shows two residuals r_1 and r_2, where each residual is sensitive to one fault of f_1 and f_2, respectively. When considering both residuals, the double-fault case $\{f_1, f_2\}$ will result in residual values in the upper right region. However, if only considering r_1, which is not sensitive to f_2, the residual output will belong to the fault mode f_1. In the same way, by considering only r_2, the residual output will belong to fault mode f_2. Thus, it is possible to classify data to belong to the double-fault mode without requiring training data from the double-fault case. However, note that the multiple-faults that can be classified using this methodology depend on the fault sensitivity of the different residuals.

In the worst case, in order to classify single-faults and double-faults, $n_f \times n_f$ 1-SVM models are required.

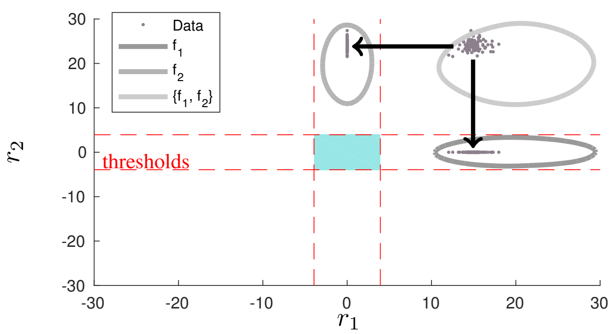

Fig. 2. If two residuals are sensitive to two different faults, f_1 and f_2, the double-fault case can be classified by classifying each single fault considering each residual separately.

C. Diagnosis system overview

Combining the model-based and data-driven fault isolation not only improves fault isolation performance but also helps to identify new unknown faults and to isolate locations in the system that they could have occurred. A summary of the diagnosis system architecture is presented here.

A schematic of the fault isolation strategy is presented in Fig. 3 where the diagnosis system is divided into two steps: fault detection and fault isolation. The fault isolation step is only activated when a fault is detected, i.e., at the event when a residual has triggered. The fault isolation step is divided into the consistency-based and the data-driven fault isolation. In the data-driven fault isolation, residual outputs are evaluated using a 1-SVM classifier for each fault mode, and is used to rank the minimal diagnosis candidates. If the residual data does not belong to any of the fault models described by the 1-SVM classifiers, it is assumed that an unknown fault has occurred. The unknown fault mode is ranked by computing the percentage of residual samples not belonging to any fault mode (8). Thus, if many of the residual values does not belong to any fault mode, it is likely that an unknown fault has occurred. Locating unknown faults can be performed, for example, using the algorithm in [9].

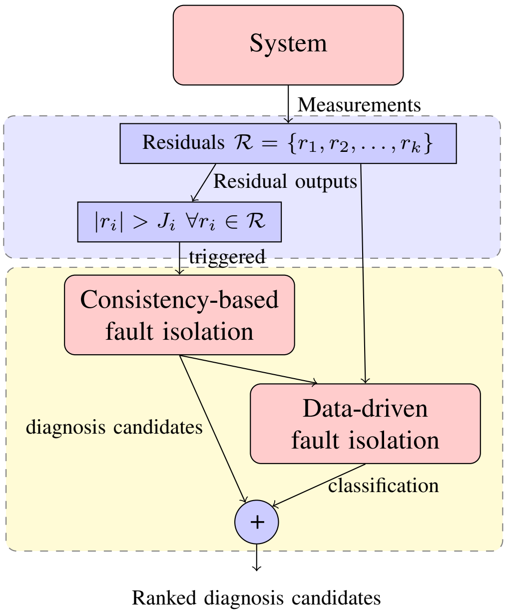

Fig. 3. A schematic of the diagnosis system design. The data-driven fault isolation is used to rank diagnosis candidates computed by the consistency-based fault isolation.

V. Evaluation

For the evaluation of the proposed diagnosis system design, a model of a generic 2 liter inline-4 cylinders single-turbocharged gasoline engine is used with 13 states describing the gas flows in the engine [5]. Three faults are considered: a clogging f_{paf} and a leakage f_{Waf} at the air filter, and a fault in the sensor measuring the intake manifold pressure f_{ypim}. The purpose of the evaluation is to illustrate the functionality of the proposed method when a pure consistency-based approach is insufficient. Before the evaluation, three 1-SVM classifier are trained using training data, i.e. one for each of the single-faults.

A. Fault isolation and ranking of diagnosis candidates

To illustrate the combined fault isolation approach, two residuals, r_1 and r_2 are selected where consistency-based single-fault isolation is limited as shown in the fault signature matrix in Table II. One solution to include the possibility of an unknown fault, is to add an extra column in the fault signature matrix to accommodate for the effect of an unknown fault f_x. The fault isolability matrix for the two residuals is given in Table III. It can be seen that f_{paf} and f_{Waf} are not isolable from each other and f_{ypim} is not isolable from the other faults.

Fig. 4 shows a 2D plot of the two residuals plotted against each other. The dashed lines represent the thresholds for each

TABLE II

Fault signature matrix for r_2 and r_3.

	f_{paf}	f_{Waf}	f_{ypim}	f_x
r_1	X	X	0	X
r_2	X	X	X	X

TABLE III

Fault isolability matrix given the two residuas

	f_{paf}	f_{Waf}	f_{ypim}
f_{paf}	X	X	
f_{Waf}	X	X	
f_{ypim}	X	X	X

residual and the solid lines represent the boundaries of the 1-SVM classifiers for each of the three single-fault modes. The 1-SVM classifier of the fault f_{paf} covers mainly the upper left region of the residual space and the model for f_{ypim} covers the area in between the thresholds for r_1 since only r_2 is sensitive to the sensor fault. Each fault mode is trained using only detected faulty data, i.e. no data is used from missed detections since those cases will not be considered when the diagnosis system is run on-line in the system. However, if the data is scattered over a large area close to the nominal residual behavior, the boundary of the 1-SVM model might overlap the nominal behavior as well. Also, although f_{Waf} should ideally trigger both residuals, most of the area belonging to its 1-SVM model corresponds only to that of r_1 being triggered. The corresponding minimal diagnosis candidates for different combinations of triggered residuals are tabulated in Table IV.

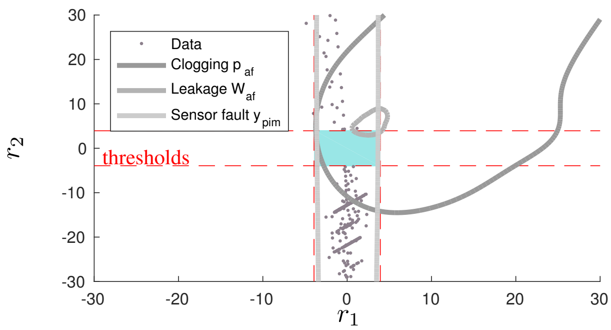

Fig. 4. Triggered residual data from a sensor fault.

TABLE IV

Computed minimal diagnosis candidates for different combinations of triggered residuals.

Triggered residuals	Minimal diagnosis candidates
no triggered residuals	$\emptyset$
r_1	$\{f_{paf}\}, \{f_{Waf}\}, \{f_x\}$
r_2	$\{f_{paf}\}, \{f_{Waf}\}, \{f_{ypim}\}, \{f_x\}$
r_1 and r_2	$\{f_{paf}\}, \{f_{Waf}\}, \{f_x\}$

In Fig. 4, evaluation data from f_{ypim} is visualized. Since only r_2 has triggered, all single-faults are minimal diagnosis candidates. The ranking of each minimal diagnosis candidate is Rank($\{f_{ypim}\}$) $= 1.00$, Rank($\{f_{paf}\}$) $= 0.11$, Rank($\{f_{Waf}\}$) $= 0.00$, and Rank($\{f_x\}$) $= 0.00$, where the true diagnosis candidate has the highest rank.

The evaluation data from the detected leakage is shown in Fig. 5. It is visible that detection performance of r_1 is not ideal since most samples lies within the thresholds. However, both residuals trigger and the ranking of the different minimal diagnosis candidates are Rank($\{f_{paf}\}$) $= 1.00$, Rank($\{f_{Waf}\}$) $= 0.77$, and Rank($\{f_x\}$) $= 0.00$. The true diagnosis candidate $\{f_{Waf}\}$ is ranked high but $\{f_{paf}\}$ is ranked higher since the boundary of $\{f_{Waf}\}$ lies within the boundary of $\{f_{paf}\}$. It is possible to improve the separation between the different faults by including other residuals.

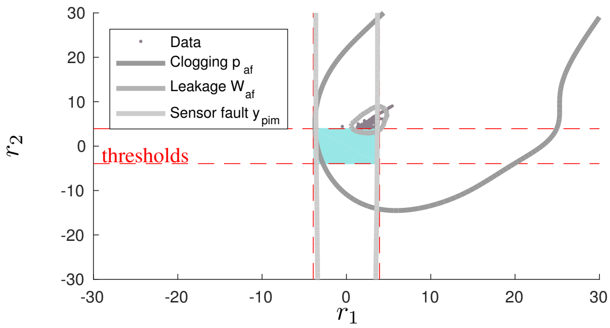

Fig. 5. Triggered residual data from a leakage.

B. Fault isolation of unknown faults

To evaluate how the algorithm would behave to an unknown fault, a new fault is simulated in the model and the results are shown in Fig. 6. The ranking of the different minimal diagnosis candidates in this scenario are Rank($\{f_{paf}\}$) $= 0.20$, Rank($\{f_{Waf}\}$) $= 0.02$, and Rank($\{f_x\}$) $= 0.80$ where the occurrence of an unknown fault has the highest rank.

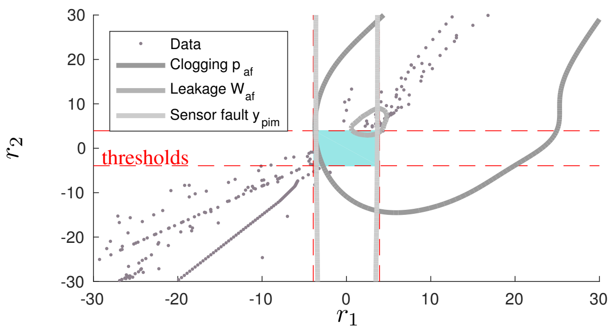

Fig. 6. Triggered residual data from an unknown fault.

VI. Conclusions

The results from the case study show the advantages of the combined consistency-based and data-driven methods for fault isolation. The data-driven fault classification improves the fault isolation and gives a priority list of the different diagnosis candidates. Since the consistency-based fault isolation and data-driven fault classification are performed sequentially, it is possible to separate the fault isolation algorithm into two different systems. Fault detection and the consistency-based fault isolation can run on-line in a system and the data-driven fault isolation can be performed during times when the system is off-line, for example, during troubleshooting. This is beneficial in automotive applications, especially when the computational capabilities in vehicles are too restrictive. If the data-driven fault isolation is performed at a workshop, it is also possible to continuously update and improve on the 1-SVM models with new residual data when faults are identified. This is a great advantage since the number of vehicles is large and data from many vehicles can be used as training data.

References

[1] C. Chen and M. Pecht. Prognostics of lithium-ion batteries using model-based and data-driven methods. In *Prognostics and System Health Management (PHM), 2012 IEEE Conference on*. IEEE, 2012.

[2] M.-O. Cordier, P. Dague, F. Levy, J. Montmain, M. Staroswiecki, and L. Trave-Massuyes. Conflicts versus analytical redundancy relations: a comparative analysis of the model based diagnosis approach from the artificial intelligence and automatic control perspectives. *SMC, Part B: Cybernetics, IEEE Transactions on*, 34(5):2163–2177, 2004.

[3] J. De Kleer and B. Williams. Diagnosing multiple faults. *Artificial intelligence*, 32(1):97–130, 1987.

[4] S. Ding, P. Zhang, T. Jeinsch, E. Ding, P. Engel, and W. Gui. A survey of the application of basic data-driven and model-based methods in process monitoring and fault diagnosis. In *Proceedings of the 18th IFAC World Congress*, pages 12380–12388, 2011.

[5] Lars Eriksson. Modeling and control of turbocharged si and di engines. *OGST-Revue de l'IFP*, 62(4):523–538, 2007.

[6] D. Jung, H. Khorasgani, E. Frisk, M. Krysander, and G. Biswas. Analysis of fault isolation assumptions when comparing model-based design approaches of diagnosis systems. *IFAC-PapersOnLine*, 48(21):1289–1296, 2015.

[7] I. Loboda and S. Yepifanov. A mixed data-driven and model based fault classification for gas turbine diagnosis. In *ASME Turbo Expo 2010: Power for Land, Sea, and Air*, pages 257–265. American Society of Mechanical Engineers, 2010.

[8] J. Luo, M. Namburu, K. Pattipati, L. Qiao, and S. Chigusa. Integrated model-based and data-driven diagnosis of automotive antilock braking systems. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 40(2):321–336, 2010.

[9] X. Pucel, W. Mayer, and M. Stumptner. Diagnosability analysis without fault models. In *20th International Workshop on Principles of Diagnosis (DX-09)*, pages 67–74, 2009.

[10] C. Sankavaram, B. Pattipati, A. Kodali, K. Pattipati, M. Azam, S. Kumar, and M. Pecht. Model-based and data-driven prognosis of automotive and electronic systems. In *Automation Science and Engineering, 2009. CASE 2009. IEEE International Conference on*, pages 96–101. IEEE, 2009.

[11] B. Schölkopf, J. Platt, J. Shawe-Taylor, A. Smola, and R. Williamson. Estimating the support of a high-dimensional distribution. *Neural computation*, 13(7):1443–1471, 2001.

[12] B. Schölkopf, R. Williamson, A. Smola, J. Shawe-Taylor, J. Platt, et al. Support vector method for novelty detection. In *NIPS*, volume 12, pages 582–588. Citeseer, 1999.

[13] N. Shashoa, G. Kvaščev, A. Marjanović, and Ž. Djurović. Sensor fault detection and isolation in a thermal power plant steam separator. *Control Engineering Practice*, 21(7):908–916, 2013.

[14] V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S. Kavuri. A review of process fault detection and diagnosis: Part i: Quantitative model-based methods. *Computers & chemical engineering*, 27(3):293–311, 2003.

[15] S. Yin, S. Ding, X. Xie, and H. Luo. A review on basic data-driven approaches for industrial process monitoring. *Industrial Electronics, IEEE Transactions on*, 61(11):6418–6428, 2014.