View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Apollo

Generating Real-valued Failure Data for Prognostics Under the Conditions of Limited Data Availability

Gishan Don Ranasinghe
Institute for Manufacturing
Department of Engineering
University of Cambridge
Cambridge, United Kingdom
Email: gd416@cam.ac.uk

Ajith Kumar Parlikad
Institute for Manufacturing
Department of Engineering
University of Cambridge
Cambridge, United Kingdom
Email: aknp2@cam.ac.uk

***Abstract*—Data-driven prognostics solutions underperform under the conditions of limited failure data availability since the number of failure data samples is insufficient for training prognostics models effectively. In order to address this problem, we present a novel methodology for generating real-valued failure data which allows training datasets to be augmented so that the number of failure data samples is increased. In contrast to existing data generation techniques which duplicate or randomly generate data, the proposed methodology is capable of generating new and realistic failure data samples. To this end, we utilised the conditional generative adversarial network and auxiliary information pertaining to the failure modes. The proposed methodology is evaluated in a real-world case study involving the prediction of air purge valve failures in heavy trucks. Two prognostics models are developed using gradient boosting machine and random forest classifiers. It is shown that when these models are trained on the augmented training dataset, they outperform the best prognostics solution previously proposed in the literature for the case study by a large margin. More specifically, costs due to breakdowns and false alarms are reduced by 44%.**

I. Introduction

Prognostics involve predicting time to failure of equipment or predicting the probability that a piece of equipment operates without a failure up to some future time [1]. Prognostics are typically random or unknown, hence they must be estimated using expert knowledge, condition monitoring data and/or event data relating to past failures. Despite their popularity, the long-lasting problem with data-driven prognostics is that they rely on large amounts of historical failure data to estimate prognostics model parameters [2]. Nevertheless, historical failure data are limited in real-world industrial scenarios [3]. This makes it difficult for data-driven models to extract degradation patterns and characterise system performance from historical data for prognostics modelling [4]. Hence, predictions produced by these models are associated with high uncertainty and therefore introduce additional costs due to under maintenance and over maintenance. The objective of this paper is therefore to propose a methodology to generate real-valued failure data, and hence augment historical datasets used for prognostics modelling to include an increased number of failure data samples. This allows predictions produced by data-driven prognostics models to be associated with minimal error and uncertainty when real failure data are limited.

One of the main reasons for limited failure data availability in industrial scenarios is the rare failures problem [3]. This problem is caused by the infrequent occurrence of failures under a single failure mode [3]. In most real-world scenarios rare failures can be disastrous [3]. The majority of industrial organisations are prepared to handle the consequences of common failures as they are experiencing them regularly. On the other hand, since rare failures are infrequent organisations have much less experience with them, hence left exposed defencelessly to suffer from the unexpected consequences of these failures. Vehicle, aircraft and telecommunications equipment maintenance are examples of industrial scenarios that constantly facing major penalties due to the rare failures problem [3], [4], [5].

When failure data are limited for data-driven prognostics, the use of physical model-based and rule-based prognostics solutions have been unsuccessful in most industrial scenarios due to following: physical model-based prognostics require the assumption or empirical estimation of physics parameters which is difficult and expensive in industrial scenarios [2]. Moreover, large amounts of historical failure data are still required for validating physical models [2]. Rule-based prognostics involve obtaining domain knowledge and converting it into rules which is also difficult in most industrial scenarios [2]. More importantly, when the number of rules increases rule-based prognostics solutions suffer from the combinatorial explosion problem [2].

Existing techniques used to address the problem of limited failure data availability for data-driven prognostics include undersampling and oversampling. Unfortunately, these techniques also have major shortcomings. Undersampling discards potentially useful non-failure data samples, hence, for instance, can degrade the discriminating power of a classifier [5]. Since oversampling techniques including advanced techniques such as the synthetic minority oversampling

Gishan Don Ranasinghe's research is funded by the Engineering and Physical Sciences Research Council (EPSRC).

978-1-5386-8357-6/19/$31.00 ©2019 IEEE

technique (SMOTE) involve duplicating existing failure data or randomly generating data, they do not introduce new and realistic (i.e. real-valued) failure data samples. Hence, the fundamental problem of limited failure data availability is not addressed [3].

We propose a methodology that overcomes the shortcomings of existing techniques by strategically generating real-valued failure data. More specifically, after identifying failure modes of the target equipment using failure mode and effect analysis (FMEA), the proposed methodology estimates a generative model that captures the semantic features of a failure mode from real failure data samples. Then it uses the generative model to generate new failure data by sampling from a joint distribution of noise and auxiliary information pertaining to the failure mode. In this methodology, the utilisation of auxiliary information available in the prognostics domain (e.g. operating environment information, expert knowledge, maintenance records, physics of failure and weather conditions) is explored for conditioning the noise being added into the newly generated data samples. The tool used for estimating the generative model is the conditional generative adversarial network (CGAN) [6]. It is an extension to the generative adversarial network (GAN) which was recently introduced as a novel way to train deep generative models in a minimax game [7]. The CGAN has been highly successful in the image recognition domain for generating real-valued images when the number of real image samples is insufficient for training image recognition models effectively [8].

Despite its success in the image recognition domain, generating real-valued data for prognostics presents the following domain-specific research challenges: (i) identifying auxiliary information pertaining to failure modes in different industrial scenarios that is useful for conditioning the noise; (ii) converting different kinds of auxiliary information that are in complex and different forms into an easily manipulated form, so that they can be integrated into the data generation process; (iii) quantifying the change in information availability for data-driven prognostics due to the generated failure data samples. The research presented in this paper takes the first steps towards overcoming these challenges, and hence developing a methodology for generating real-valued failure data for data-driven prognostics under the conditions of limited failure data availability.

The remainder of the paper is organised as follows: Sec. II summarises relevant aspects of the theoretical background of CGAN. Sec. III presents the mathematical formulation of the problem of prognostics under the conditions of limited failure data availability. The proposed methodology for generating real-valued failure data is discussed in Sec. IV. The methodology is evaluated in the Scania air purge valve prognostics problem and results are presented in Sec. V. The paper is concluded and future work is outlined in Sec. VI.

II. Background

A. Generative Adversarial Network

The standard GAN can be used to generate data as follows. First, a generative model is trained in an adversarial training framework to generate data similar to the real data distribution. The adversarial training framework allows a model to estimate its parameters by competing with another model (i.e. two-player minimax game) [7]. Then new data samples are generated by using the generative model to sample data from a data distribution containing random noise [7].

The GAN consists of two neural networks: a generator G and a discriminator D. Given a dataset X with samples $\{x \in X\}$, the goal of a GAN is to estimate a generative model that captures the generator distribution $P_G(x)$ that matches the real data distribution $P_{\text{data}}(x)$. The GAN estimates this generative model by first enabling it to sample data from P_G by transforming a prior noise variable $z \sim P_{\text{noise}}(z)$ into a new data sample $G(z)$. Then the discriminator network D is used to discriminate between whether the generated sample $G(z)$ is a real data sample (i.e. $G(z)$ is sampled from the real data distribution P_{data}) or a fake data sample (i.e. $G(z)$ is sampled from the generator distribution P_G). Thus, the discriminator outputs a single scaler indicating the probability of whether a given data sample is real or fake without knowing the sample is generated by the generator. The generator G uses this scaler as the feedback to minimise its loss function $\log(1-D(G(z)))$ whilst the discriminator D tries to minimise the loss function $\log(D(x))$ to improve its discriminating power. Formally, the value function $V(G, D)$ of the minimax game for estimating a generative model using the GAN is as follows:

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{\text{data}}}[\log D(x)] + \mathbb{E}_{z\sim\text{noise}}[\log(1-D(G(z)))] \tag{1}$$

B. Conditional Generative Adversarial Network

In the standard GAN, the generative model is trained without conditioning the noise being added into generated data samples [6]. Thus, the generation of real-valued data is not guaranteed, since there is no control over the modes of data being generated [6]. In the image recognition domain, this issue is addressed by extending the GAN into the CGAN [6]. In contrast to the GAN, CGAN conditions the generator and discriminator on auxiliary information [6]. This is done by feeding a vector representation of auxiliary information into the generator and discriminator neural networks as additional inputs.

More specifically, using auxiliary information vector Y with samples $\{y \in Y\}$, the generator G in CGAN is modified to generate data samples $G(z|y)$ compared to $G(z)$ in the GAN. This means G generates a fake data sample $G(z|y)$ from the joint distribution of noise and auxiliary information $P(Z,Y)$. Similarly, the discriminator is extended to $D(x|y)$ compared to $D(x)$ in the GAN. Thus, the discriminator tries to discriminate between real and fake data samples by detecting whether a given sample is sampled from the joint distribution of real data and auxiliary information $P(X,Y)$. The feedback from the discriminator allows the generator to condition noise on auxiliary information since it now needs to generate fake data samples that fool the discriminator in two cases: (i) when discriminating against real data samples; (ii) when discriminating against other information related to the prediction task.

III. Problem Formulation

Before formulating the problem we introduce a method for measuring the limited failure data availability. Prognostics datasets become imbalanced when there is a limited number of failure data samples [3]. This means there is a relatively large number of non-failure data samples compared to the failure data samples. The balance of a dataset can be measured using the Shannon entropy, which calculates the average rate at which the information is produced by a stochastic dataset [9]. Formally, for a dataset with N number of total data samples and K number of classes with size C_1 to C_K, the normalised Shannon entropy H' is given by (2). When H' gets closer to 0 the extent of the limited failure data availability problem increases and conversely, when it gets closer to 1 the extent of the problem decreases.

$$H' = \frac{-\sum_{i=1}^{K} \frac{C_i}{N} \log \frac{C_i}{N}}{\log K} \tag{2}$$

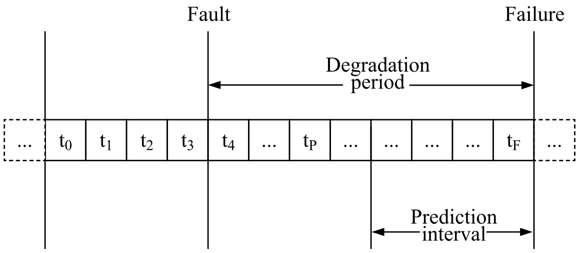


Fig. 1. Diagram depicting key terms used for the problem formulation.

The problem of prognostics under the conditions of limited failure data availability can be formulated using Fig. 1. When a piece of equipment develops a fault, the fault grows in magnitude causing a monotonic or nonmonotonic degradation until failure. In order to predict the failure, starting from time t_P we observe condition monitoring data and/or event data $X_t \in \mathbb{R}, t > t_P$. The data before time t_P is known, that is, we have a historical training dataset $D = \{X_t\}_{t=1}^{t_P}$. C_1 is a conditional statement which specifies the ideal prediction interval for prognostics. It indicates how long before the failure the prognostics model needs to predict it. The prediction interval is determined with the help of a domain expert.

More importantly, the training dataset D has a limited number of failure data samples. Hence, D satisfies the following conditional statement $C_2 : 0 \leq H' \leq L$, where H' is the normalised Shannon entropy of D and L is the highest value of H' in which the existing prognostics solutions start to underperform in an industrial scenario due to the limited failure data availability.

The basic task is to learn a prediction procedure P_1 that predicts the possible future failures with minimal error and uncertainty using the training dataset D whilst satisfying C_1. Thus, P_1 is a function that maps a failure event sequence S in D to a boolean prediction value which indicates whether there will be a future failure. Hence, $P_1 : S \rightarrow \{0,1\}$.

Using reliability theory P_1 can be extended to the time to failure prediction of equipment procedure P_2 as follows: calculate $\eta.\Gamma(\frac{1}{\beta}+1)$ where η and β are the scale and shape parameters of the Weibull distribution, and Γ is the gamma function. Note that $\eta.\Gamma(\frac{1}{\beta}+1)$ is the formulae for calculating mean time to failure. Thus, the prediction procedure P_2 is to estimate η and β using failure event sequence S.

Using probability theory P_1 can be extended to predicting the probability that a piece of equipment operates without a failure up to some future time prediction procedure P_2 as follows: calculate $P(X|Y)$, meaning for a given data sample $x \in X$ assign the probability $y \in Y$, where y is a label that indicates whether the given data sample x contains a degradation pattern pertaining to the failure mode that needs predicting. Thus, the prediction procedure P_3 is to estimate $P(X|Y)$ using S and a classifier $f(x)$ for the time period $t_P + W$, where W is the future time window that includes the actual time of failure t_F. This window is determined by the minimum lead time required to plan and schedule maintenance tasks, deploy maintenance engineers and perform maintenance.

One can observe that prediction procedure P_1 is already satisfied by either P_2 or P_3 (i.e. P_2 and P_3 already predict whether there will be a future failure). Thus, the objective of prognostics under the conditions of limited failure data availability is to perform prediction procedures P_2 and P_3 when training dataset D satisfies the conditional statement relating to the limited failure data availability C_2.

IV. Proposed Methodology

Fig. 2 shows a flowchart of the proposed methodology for generating real-valued failure data in industrial scenarios that face the problem of limited failure data availability for prognostics modelling.

A. Phase 1: Identify the Failure Mode and Baseline Prognostics Performance

1) Perform FMEA or use expert knowledge to identify the failure mode of the target equipment: FMEA is performed to identify the failure mode of the target equipment in order to get an understanding of what condition monitoring and/or event data need to be captured for prognostics. One can also use the literature for this purpose since for the majority of industrial equipment failure modes are already known and presented in the existing literature.

2) Obtain a historical dataset consists of condition monitoring and/or event data pertaining to the failure mode: Once the failure mode is identified, condition monitoring and/or event data pertaining to the failure mode are captured from the target equipment. Condition monitoring data include measurements related to the health condition of the equipment.

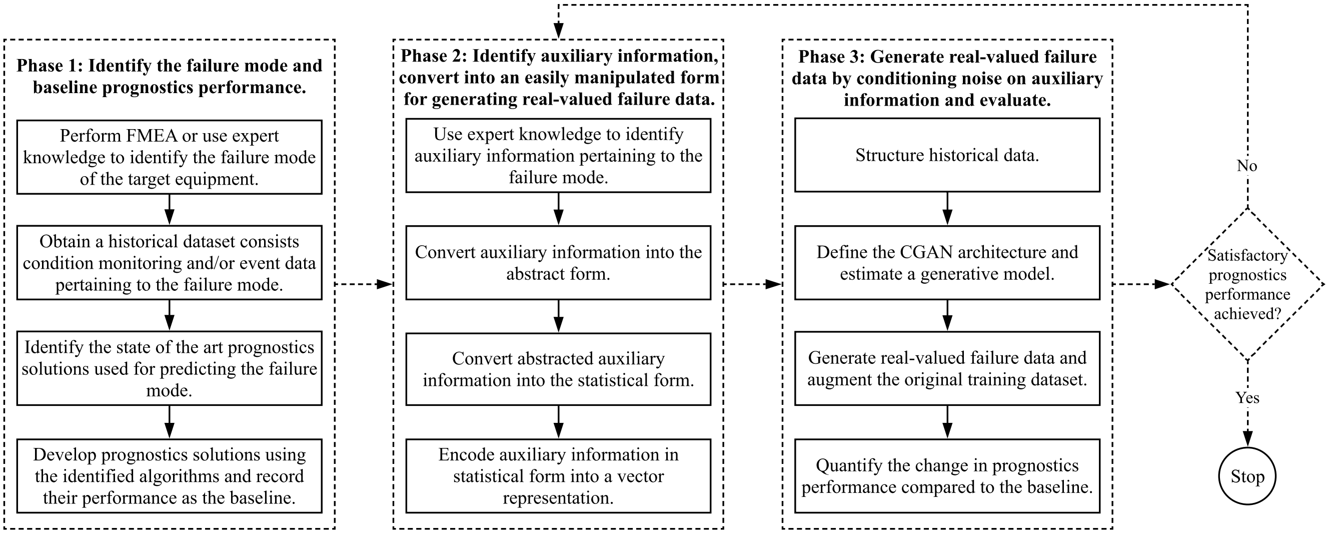

Phase 1: Identify the failure mode and baseline prognostics performance.
Perform FMEA or use expert knowledge to identify the failure mode of the target equipment.
Obtain a historical dataset consists condition monitoring and/or event data pertaining to the failure mode.
Identify the state of the art prognostics solutions used for predicting the failure mode.
Develop prognostics solutions using the identified algorithms and record their performance as the baseline.
Phase 2: Identify auxiliary information, convert into an easily manipulated form for generating real-valued failure data.
Use expert knowledge to identify auxiliary information pertaining to the failure mode.
Convert auxiliary information into the abstract form.
Convert abstracted auxiliary information into the statistical form.
Encode auxiliary information in statistical form into a vector representation.
Phase 3: Generate real-valued failure data by conditioning noise on auxiliary information and evaluate.
Structure historical data.
Define the CGAN architecture and estimate a generative model.
Generate real-valued failure data and augment the original training dataset.
Quantify the change in prognostics performance compared to the baseline.
Satisfactory prognostics performance achieved?
No
Yes
Stop

Fig. 2. Flowchart of the proposed methodology for generating real-valued failure data for prognostics under the conditions of limited failure data availability.

They can be captured from supervisory control and data acquisition (SCADA) systems or deploying internet of things sensors. Event data include information related to previous failures of the equipment which can be captured from inspection records, maintenance work-orders, repair and replacements records.

3) Identify the state of the art prognostics solutions used for predicting the failure mode and record their performance as the baseline: If previous researchers have already attempted to develop prognostics solutions for predicting failure of the target equipment under the identified failure mode, the best set of solutions currently available are used as the baseline. If no previous work has been done, expert knowledge and the process of elimination is used to identify the state of the art prognostics solutions from a set of solutions implemented using various statistical and machine learning algorithms. The best performance the state of the art solutions can achieve on the historical dataset with a limited number of failure data samples is recorded as the baseline prognostics performance.

The prognostics performance is measured using precision and recall. The standard evaluation metrics such as accuracy and error rate are not suitable for evaluating prognostics models when failure data are limited since they will be biased to the negative class (i.e. non-failure data class) regardless of the positive class (i.e. failure data class) leads to the poor performance [10]. Precision and recall, however, are not affected by the majority class [10]. The precision is the fraction of correctly predicted failures among all the predicted instances that include actual failures and false alarms. The recall is the fraction of correctly predicted failures among all the actual failures. This means higher the precision lower the false alarms (i.e. lower the false positive rate), and higher the recall lower the undetected failures (i.e. lower the false negative rate).

B. Phase 2: Identify Auxiliary Information and Convert into an Easily Manipulated Form for Generating Real-valued Failure Data

1) Use expert knowledge to identify auxiliary information pertaining to the failure mode: Table I outlines different kinds of auxiliary information available in the prognostics domain. The challenge is to identify pieces of auxiliary information that are useful for generating real-valued failure data in different industrial scenarios. In the proposed methodology, expert knowledge provided in the literature and obtained from on-site maintenance engineers is used to identify auxiliary information that may potentially be useful for generating real-valued failure data. Then phase two and three are iteratively performed in order to identify the ideal set of auxiliary information that leads to the satisfactory prognostics performance.

TABLE I
DIFFERENT KINDS OF AUXILIARY INFORMATION AVAILABLE IN THE PROGNOSTICS DOMAIN

Sources of Epistemic Uncertainty	**Expert Knowledge**
Operating environment information Equipment stress level information Weather conditions	Equipment similarity information Empirically validated rules Known failure thresholds
Maintenance Text Records	**Physics of Failure**
Inspection records Repair and replacement records	Differential equations Stochastic differential equations

2) Convert auxiliary information into an easily manipulated form for generating real-valued failure data: Auxiliary information pertaining to the failure modes is in complex and different forms. For example, text entries of maintenance records, differential equations of physics of failure and meteorological data representing weather conditions. Thus, the challenge is to convert this information into an easily manipulated form in order to enable them to be integrated into the data generation process for conditioning the noise.

In the proposed methodology, auxiliary information is converted into vector representations. This can be further explained using the following example. Imagine that there is a set of equipment that has failed under the same failure mode. During the past degradation periods of this set of equipment, the maintenance engineers have taken measurements of their stress levels at regular intervals. We can use this stress level information to generate real-valued failure data by informing the CGAN that a newly generated failure data sample may contain the patterns in historical stress level data the equipment had during their past degradation periods. Thus, the noise being added to the generated data samples is conditioned on stress level information related to the past failures of the equipment.

In order to integrate stress level information into the data generation process, we first convert it into an abstract form. This allows equipment-specific information to be generalised to all the equipment that has failed under the failure mode that needs predicting. For instance, if the stress level information of equipment is recorded as *the surface temperature of equipment A, B and C increased from 40 to 80 Celsius*, once converted into the abstract form this information becomes *some variable X increases*. Thus, specific terms such as equipment A, B and C, surface temperature and numerical thresholds are removed. Then the abstracted information is converted into the statistical form by representing it as some continuous variable C. The continuous variable C can be converted into a distribution between some values y_0 and y_1. Finally, this distribution can be represented as a vector Y containing some values $\{y \in Y | y_0 < y < y_1, \text{and y increases}\}$.

C. Phase 3: Generate Real-valued Failure Data by Conditioning Noise on Auxiliary Information and Evaluate

1) Structure historical data: As shown in Fig. 3, the historical data obtained in phase one is divided into three datasets: (i) training dataset (referred to as the original training dataset) which includes data for training prognostics models; (ii) validation dataset which is used for hyperparameter tuning; (iii) testing dataset which is used to evaluate prognostics models on previously unseen data.

The objective of generating real-valued failure data is to augment the original training dataset so that the number of failure samples available for training prognostics models is increased. To this end, as shown in Fig. 3 we first divide the original training dataset into two subsets containing failure data (referred to as the training failure data subset) and non-failure data (referred to as the training non-failure data subset). The training failure data subset is used to estimate a generative model that captures the semantic features of the failure mode using noise and auxiliary information vectors. After the dataset containing real-valued failure samples is generated, it is combined with the two subsets to obtain the augmented training dataset. The validation and testing datasets are left unchanged for hyperparameter optimisation and comparative evaluation of prognostics models.

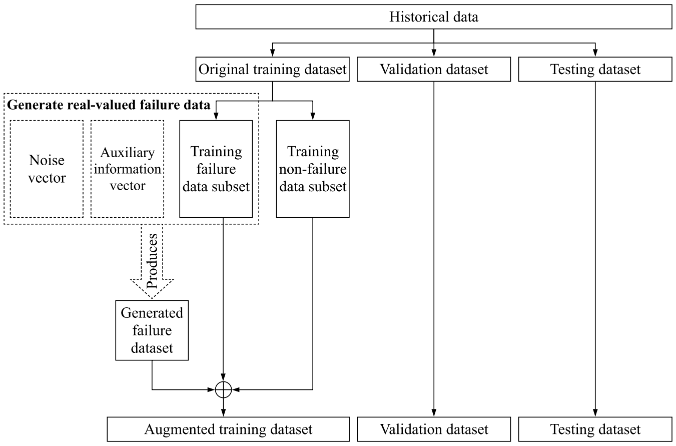

Historical data
Original training dataset
Validation dataset
Testing dataset
Generate real-valued failure data
Noise vector
Auxiliary information vector
Training failure data subset
Training non-failure data subset
Produces
Generated failure dataset
Augmented training dataset
Validation dataset
Testing dataset

Fig. 3. Diagram depicting how historical data are structured in the proposed methodology. The original training dataset is augmented by integrating generated real-valued failure data samples. The validation and testing datasets are left unchanged for hyperparameter optimisation and comparative evaluation of prognostics models.

2) Define the CGAN architecture and estimate a generative model: Following from the theoretical background discussed in Sec. II, the CGAN architecture implemented for our methodology is presented in Fig. 4. In the proposed methodology, the generator and discriminator are two artificial neural networks (ANN). As denoted by number *1* in the figure, first the noise vector Z is combined with the auxiliary information vector Y into the joint distribution $P(Z, Y)$. This is then used as the input to the generator ANN. Then as denoted by the number *2*, data samples in the training failure data subset X are combined with the auxiliary information vector Y into the joint distribution $P(X, Y)$. This is used as the input to the discriminator ANN.

The objective of the generator G is to fool the discriminator D into believing that a generated failure data sample is real. Thus, as denoted by the number *3* in Fig. 4, the generator produces a generated data sample $G(z|y)$ by conditioning the noise on auxiliary information. More specifically, the generator aims to minimise its loss function $\log(1 - D(G(z|y)))$. The objective of the discriminator D is to detect whether a given data sample is real or fake. Thus, as denoted by the number *4*, the discriminator produces a probability $D(x\prime|y)$ indicating how much it believes the given data sample x' is real or fake. More specifically, the discriminator tries to minimise its loss function $\log(D(x'|y))$. At the end of the training, the generator is capable of generating new failure data samples that the discriminator cannot discriminate as real or fake. Thus, the generator ANN (i.e. estimated generative model) has now captured the semantic features of the failure mode and hence can be used for generating real-valued failure data.

3) Generate real-valued failure data and augment the original training dataset: A new noise vector and the previously used auxiliary information vector are used as inputs to the estimated generative model for generating real-valued failure data. More specifically, given the joint distribution of

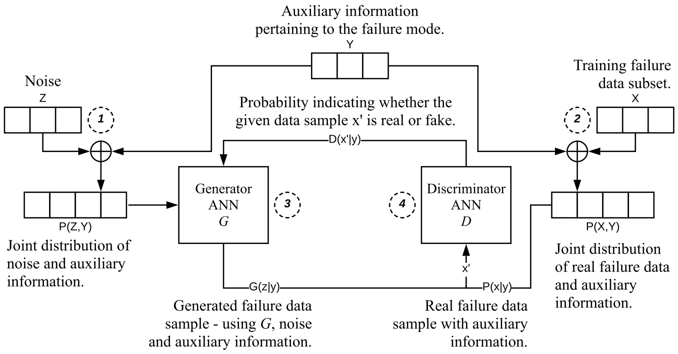


Fig. 4. Diagram depicting the architecture of the conditional generative adversarial network implemented for the proposed methodology.

noise and auxiliary information vectors as input, the generative model can predict a set of real-valued failure data samples. Once the real-valued failure data samples are generated, they are combined with the original training dataset to obtain the augmented training dataset as shown in Fig. 3.

4) Quantify the change in prognostics performance compared to the baseline: The change in performance is quantified by comparing the baseline precision and recall to the precision and recall obtained when prognostics models are trained on the augmented training dataset and evaluated on the testing dataset.

V. Scania Case Study

A. The Problem of Scania Air Purge Valve Prognostics

Scania heavy trucks are popular for their customisability since the customers are provided with the ability to choose from a wide range of options for customising trucks to match to their specific requirements [11]. The number of axels, power take-off position, engine power, fuel type are examples of customisable options. Moreover, Scania trucks are used for different purposes (e.g. long haulage, construction work and garbage collection) [11]. This diversity presents a challenge for the planned maintenance of trucks, hence a prognostics solution that predicts failures in individual trucks with minimal error and uncertainty is required [11].

The air dryer is part of the air processing system that provides compressed air for the critical components of heavy trucks such as air brakes, air suspension and gearbox. It removes water vapour from compressed air to prevent moisture and condensation from interfering with critical components. This process of removing water vapour from compressed air is called air purging. Air dryer uses an air purge valve (APV) (see Fig. 5) to automatically purge compressed air. These valves can degrade in performance due to crack, hence cause compressed air to be leaked from the air processing system. APV failures often result in the complete immobilisation of vehicles since there is insufficient amount of compressed air in the truck for performing its critical functions [11].

Fig. 5. Image of an air dryer with the air purge valve marked in red.

Scania provided a dataset that contains data collected from 80000 heavy trucks from five European markets [11]. The positive class in the dataset represents trucks with APV failures. The prognostics problem is modelled as a binary classification task in which the challenge is to predict whether a truck faces an APV failure in near future. According to Scania, the estimated cost of an undetected APV failure is €500 (C_{FN}), and the estimated cost of a false alarm is €10 (C_{FP}) [11]. C_{FN} is due to the cost of an undetected APV failure that causes a breakdown of a truck. C_{FP} is due to the unnecessary checks that need to be done by a mechanic due to false alarms. The objective is to reduce the total cost of breakdowns and false alarms. Let m be the number of undetected APV failures and n be the number of false alarms, then the total cost of breakdowns and false alarms T_{Cost} is given by the following:

$$T_{\text{Cost}} = mC_{FN} + nC_{FP} \tag{3}$$

The correct prediction of APV failures needs to be given priority over the false alarms since the undetected failures result in a larger penalty (i.e. €500 compared to €10). Hence, as shown in (4) using the prediction procedure $P3$ defined in Sec. III, we formulate the Scania APV prognostics problem as an optimisation problem. Thus, the objective of the problem of Scania APV prognostics is to minimise T_{Cost} whilst optimising P_3 to predict the optimal value pair for m and n with minimal error and uncertainty. Hereinafter, we refer to T_{Cost} to as the *prognostics performance*.

$$\begin{aligned} \underset{P_3}{\text{minimise}} \quad & T_{Cost} = mC_{FN} + nC_{FP} \\ \text{subjected to} \quad & m \geq 0, \\ & n \geq 0, \\ & C_{FN} = 500, \\ & C_{FP} = 10. \end{aligned} \tag{4}$$

B. Limited Failure Data Availability for Scania APV Prognostics

Scania dataset is divided into training and testing datasets. The training dataset contains 60000 data samples and the testing dataset contains 16000 data samples. Out of the 60000 training samples only 1000 belongs to the positive class. This imbalance ratio of 1000:59000 between positive and negative classes means that the positive class only covers 1.6% of the entire training dataset, whereas the negative class covers 98.4%. Thus, the Scania dataset is considered as a highly imbalanced dataset in the literature [12].

Using the method introduced in Sec. III, we measure the extent of the problem of limited failure data availability. In this case, the number of positive samples C_1 and negative samples C_2 are 1000 and 59000 respectively. The number of classes K is 2. The normalised Shannon entropy H' of the Scania dataset is therefore 0.08 which indicates a highly imbalanced dataset, thus the extent of the limited failure data availability problem is high.

C. The State of the Art of Scania APV Prognostics

There are a few solutions already proposed in the literature for addressing the problem of Scania APV prognostics. Table II summarises the top three prognostics solutions currently proposed in the literature. Except the solution proposed in [11], other two solutions were developed during the 15th Intelligent Data Analysis (IDA 2016) competition. The testing dataset was not provided to the authors of [12] and [13] during the competition, and hence the results presented within these two publications are the prognostics performance obtained when the solutions were evaluated on the training dataset. However, the publishers of Scania dataset later published the prognostics performance of these two solutions when they were evaluated on the testing dataset [11]. Thus, the results summarised in Table II are the performance achieved when all the three solutions were evaluated on the testing dataset.

In the remainder of this section, we show that using the proposed methodology for generating real-valued failure data for prognostics under the conditions of limited failure data availability, one can obtain a far better prognostics performance than all the existing solutions.

TABLE II
PROGNOSTICS PERFORMANCE (T_{COST}) OBTAINED BY TOP THREE SOLUTIONS AVAILABLE IN LITERATURE FOR SCANIA APV PROGNOSTICS

Rank	$T_{\textbf{Cost}}$ (€)	**Undetected Failures**	**False Alarms**	**Solution Reference**	**Performance Reference**
1	9920	9	542	[12]	[11]
2	10900	12	490	[13]	[11]
3	11430	12	543	[11]	[11]

D. Generating Real-valued Failure Data for Scania APV Prognostics

In this section, a discussion on how the three phases of the proposed methodology are applied to address the problem of limited failure data availability in Scania industrial scenario is presented.

1) Phase 1: Identify the failure mode and baseline prognostics performance: Since the failure mode is previously identified and a dataset is already provided by Scania, we start with identifying the state of the art algorithms used for prognostics modelling in the Scania dataset. The random forest (RF) classifier-based prognostics solutions have been previously successful in predicting Scania APV failures (e.g. [11] and [12]). The gradient boosting machine (GBM) is another popular classifier for developing classification-based prognostics solutions [14]. We implemented two prognostics solutions using the GBM and RF classifiers. When trained on the original training dataset and evaluated on the testing dataset, these solutions have obtained prognostics performance T_{Cost} of €10750 and €11090 respectively.

The baseline performance used in this work is the prognostics performance achieved by rank one in Table II (i.e. T_{Cost} of €9920). This is the best performance achieved in the previous literature for the Scania APV prognostics [11].

2) Identify auxiliary information and convert into an easily manipulated form for generating real-valued failure data: Similarity analysis is used in the prognostics research for predicting failures in asset fleets by allowing similar assets to share data between each other [15]. In our work, similarity analysis is employed for a different purpose. We group similar trucks with APV failures in the training dataset and use these groups as auxiliary information to direct the data generation process.

Since no information that can be used to group trucks (e.g. mileage, purpose, etc.) is provided with the Scania dataset, we use clustering to identify natural groupings of trucks with APV failures. We first create a subset D' that contains only the failure data samples in the training dataset (i.e. training failure data subset). Then k-means and hierarchical clustering algorithms are used to identify the natural groupings in D'. Fig. 6 shows the t-SNE projection of the clusters generated by k-means and hierarchical clustering algorithms. It can be observed that all the data samples (i.e. trucks with APV failures) in the training failure data subset D' are grouped into two distinct clusters by both clustering algorithms. Thus, the following abstract piece of auxiliary information is identified: *there are two groups of trucks with APV failures*. Hence, the data generation process can be directed using the following condition: *a newly generated data sample representing a truck with APV failure should belong to one of two groups*.

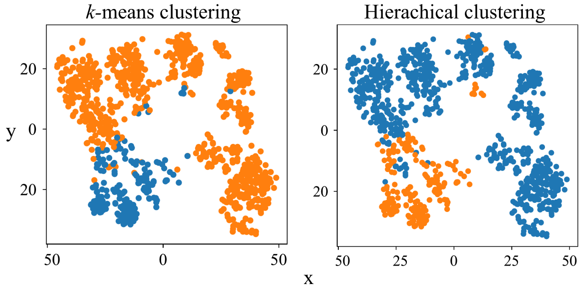


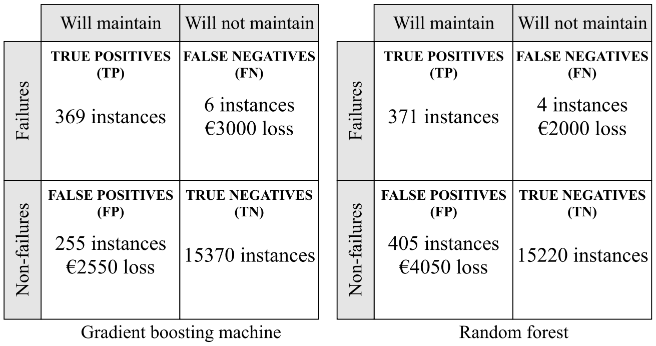
Fig. 6. The t-SNE projection of natural groupings of trucks with APV failures. It can be observed there are two groups of trucks with APV failures in the training dataset.

In order to convert abstracted auxiliary information into a vector representation, we choose the class labels generated by the k-means algorithm since it has obtained the best silhouette score which indicates the quality of clustering. The class labels that represent the two groups with natural numbers 1 and 2 is a vector of natural numbers $Y = \{y \in \mathbb{N} | 1 \leq y \leq 2\}$.

3) Phase 3: Generate real-valued failure data by conditioning noise on auxiliary information and evaluate: In order to estimate a generative model that captures the semantic features of APV failures in Scania heavy trucks, we use the CGAN architecture previously presented in Fig. 4. The generator G and discriminator D are artificial neural networks. The auxiliary information vector Y is the vector representation of class labels obtained in the previous phase. The training failure data subset X is D'. The noise vector Z is Gaussian noise. Using these parameters as inputs to the CGAN architecture, we estimate the generative model for generating real-valued APV failure data by conditioning noise on auxiliary information.

Then the original training dataset is augmented to include the generated real-valued failure data samples. In this instance, we generated 2000 failure data samples and therefore the positive and negative sample ratio in the augmented training dataset is 3000:59000 compared to the 1000:59000 in the original training dataset. Moreover, the normalised Shannon entropy H' is now increased from 0.08 to 0.2.

Fig. 7 shows reliability-based confusion matrixes obtained for GBM and RF classifier-based prognostics solutions when trained on the augmented training dataset and evaluated on the testing dataset. The prognostics performance T_{Cost} achieved by the GBM and RF classifier-based prognostics solutions are €5550 and €6050 respectively. Compared to the performance obtained by the best prognostics solution previously proposed in the literature (the baseline), this is a 44% (GBM) and 39% (RF) reduction of costs due to breakdowns and false alarms.



	Will maintain	Will not maintain
Failures	TRUE POSITIVES (TP) 369 instances	FALSE NEGATIVES (FN) 6 instances €3000 loss
Non-failures	FALSE POSITIVES (FP) 255 instances €2550 loss	TRUE NEGATIVES (TN) 15370 instances

Gradient boosting machine

	Will maintain	Will not maintain
Failures	TRUE POSITIVES (TP) 371 instances	FALSE NEGATIVES (FN) 4 instances €2000 loss
Non-failures	FALSE POSITIVES (FP) 405 instances €4050 loss	TRUE NEGATIVES (TN) 15220 instances

Random forest

Fig. 7. Reliability-based confusion matrixes depicting the prognostics performance of two prognostics solutions when trained on the augmented training dataset and evaluated on the testing dataset.

VI. Conclusion and Future Work

In this paper, the research work conducted by the authors for taking first steps towards developing a methodology that generates real-valued failure data for prognostics under the conditions of limited failure data availability is presented. This methodology integrates the conditional generative adversarial network, existing failure data, noise and auxiliary information pertaining to the failure modes for generating new and realistic failure data samples. Thus, allows predictions produced by data-driven prognostics solutions to be associated with minimal error and uncertainty when real failure data are limited.

We intend to further develop the initial version of the methodology proposed in this paper. More specifically, future research involves addressing the following challenges: (i) adapting the methodology into industrial scenarios with other kinds of historical data; (ii) identifying other kinds of auxiliary information, and how to convert and integrate them into the real-valued failure data generation process; (iii) developing an application criteria that outlines in industrial scenarios with what characteristics the methodology is suitable to be applied.

Acknowledgment

The authors would like to thank Jonas Biteus (Scania CV AB) and Tony Lindgren (Stockholm University) for disclosing the Scania Trucks dataset for public use.

References

[1] A. K. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical systems and signal processing*, vol. 20, no. 7, pp. 1483–1510, 2006.

[2] Y. Peng, M. Dong, and M. J. Zuo, "Current status of machine prognostics in condition-based maintenance: a review," *The International Journal of Advanced Manufacturing Technology*, vol. 50, no. 1-4, pp. 297–313, 2010.

[3] G. M. Weiss, "Mining with rarity: a unifying framework," *ACM Sigkdd Explorations Newsletter*, vol. 6, no. 1, pp. 7–19, 2004.

[4] Y. Zhang, G. W. Gantt, M. J. Rychlinski, R. M. Edwards, J. J. Correia, and C. E. Wolf, "Connected vehicle diagnostics and prognostics, concept, and initial practice," *IEEE Transactions on Reliability*, vol. 58, no. 2, pp. 286–294, 2009.

[5] S. Alestra, C. Brand, E. Burnaev, P. Erofeev, A. Papanov, C. Bordry, and C. Silveira-Freixo, "Rare event anticipation and degradation trending for aircraft predictive maintenance," in *11th World Congress on Computational Mechanics, WCCM*, 2014, pp. 6571–6582.

[6] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.

[7] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[8] S. Gurumurthy, R. Kiran Sarvadevabhatla, and R. Venkatesh Babu, "Deligan: Generative adversarial networks for diverse and limited data," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 166–174.

[9] J. Van Hulse, T. M. Khoshgoftaar, A. Napolitano, and R. Wald, "Feature selection with high-dimensional imbalanced data," in *2009 IEEE International Conference on Data Mining Workshops*. IEEE, 2009, pp. 507–514.

[10] S. M. A. Elrahman and A. Abraham, "A review of class imbalance problem," *Journal of Network and Innovative Computing*, vol. 1, no. 2013, pp. 332–340, 2013.

[11] J. Biteus and T. Lindgren, "Planning flexible maintenance for heavy trucks using machine learning models, constraint programming, and route optimization," *SAE International Journal of Materials and Manufacturing*, vol. 10, no. 2017-01-0237, pp. 306–315, 2017.

[12] C. F. Costa and M. A. Nascimento, "Ida 2016 industrial challenge: Using machine learning for predicting failures," in *International Symposium on Intelligent Data Analysis*. Springer, 2016, pp. 381–386.

[13] C. Gondek, D. Hafner, and O. R. Sampson, "Prediction of failures in the air pressure system of scania trucks using a random forest and feature engineering," in *International Symposium on Intelligent Data Analysis*. Springer, 2016, pp. 398–402.

[14] Z. Wu, W. Lin, and Y. Ji, "An integrated ensemble learning model for imbalanced fault diagnostics and prognostics," *IEEE Access*, vol. 6, pp. 8394–8402, 2018.

[15] A. S. Palau, K. Bakliwal, M. H. Dhada, T. Pearce, and A. K. Parlikad, "Recurrent neural networks for real-time distributed collaborative prognostics," in *2018 IEEE International Conference on Prognostics and Health Management (ICPHM)*. IEEE, 2018, pp. 1–8.