https://doi.org/10.53656/ped21-6s.20tec *Marine Engineering Education*

TECHNICAL DIAGNOSTICS OF MARINE EQUIPMENT WITH PSEUDO-DISCRETE FEATURES

Guixin Fan[1], Natalia Nikolova[1),2)], Ty Smith[1], Kiril Tenekedjiev[1), 2)]
[1]University of Tasmania (Australia)
[2]Nikola Vaptsarov Naval Academy (Bulgaria)

Abstract. We present a system for technical diagnostics (TD) that can recognize the actual state of marine equipment. A Bayesian classifier is trained to identify the different classes of a piece of equipment, monitored through multiple pseudo-discrete features. Data learning samples can be acquired with direct experiments for each class. The system is capable of merging subjective expert knowledge and data learning samples using pseudo-Bayesian estimates when the parameters of the conditional likelihood for the classes are identified. In the training process, correction is applied to solve numerical problems arising from zero probabilities. The pseudo-discrete features have hybrid nature and combine probabilistic and fuzzy approaches. They combine the ease of extracting subjective expert knowledge typical for the discrete features with the high precision of using the measured data during recognition typical for the continuous features. The domain of each pseudo discrete feature is divided into several main categories of non-overlapping intervals which are described as words by the expert. If a measured feature falls between two consecutive categories it is treated as a linear combination of those categories. The resubstitution performance of the classifier is assessed using an error matrix. A numerical example of a marine diesel generator demonstrates the proposed algorithm in a classification problem with nine different state classes of the generator, monitored through 23 pseudo-discrete features. Data learning samples are acquired with direct experiments for each class. The created TD system has potential applications in other complex engineering systems and may support improvements in marine engineering education and training.

Keywords: fuzzy-probabilistic merging; pseudo-bayesian parameter estimation; learning information, pattern recognition

Introduction

The technical diagnostics (TD) process has been a topic rising in popularity, as industry continues to seek ways to lower spendings on maintenance and minimize system downtime losses. With the aid of reliability engineering, the process of recognizing the working status of any machine or complex system into classes (Koc et al. 2012) has been simplified immensely. However, with a complex system or complex piece of machinery, the diagnostic process remains tedious for maintenance personnel. More and more unexperienced personnel, or staff unfamiliar with the workings/operations of such systems become in charge of maintenance and monitoring. Without a good knowledge of the proper workings of such systems, diagnosis of potential faults becomes problematic, since they often face large amount of monitored data without knowing the meaning behind it, or what it indicates.

In this paper, we introduce a TD system using a multi-class classifier, based on pseudo-discrete features (see (Duda et al. 2001) and (Nikolova et al. 2019) for further discussion on pseudo-discrete features). We shall demonstrate the application of the system to recognize the actual state of a hypothetical marine diesel generator in a numerical example. We shall train a Bayesian classifier to identify nine different state classes of the equipment, monitored through 23 pseudo-discrete features. For the learning process, we shall combine subjective expert knowledge and data learning samples using pseudo-Bayesian estimates (Skaggs et al. 1989) when the parameters of the conditional likelihood for the classes are identified. We shall apply epsilon correction in the training process to solve numerical problems arising from zero probabilities.

In what follows, we first present the structure of the technical diagnostics system, as well as the structure of the information we shall utilize for the training of the classifier and the epsilon corrections we shall apply. Then we present the application of the Bayesian classifier to a numerical example for the hypothetical MTU 8V396 marine diesel generator. The last section concludes the paper.

System Description

Let us analyse a complex system or a system component with several working statuses, categorized into classes. To monitor the system's working status (or class), a vector $\vec{X}$ is introduced as a d-dimensional measurement of properties (e.g. temperature, pressure, flowrate, displacement, etc.), where each measured property is a pseudo-discrete feature, represented with x_i, $i = 1,2,3...d$, as follows: $\vec{X} = \left(x_1, x_2, \dots, x_d\right)^T$

Therefore, at a given measured property of the component, the posterior probability of the component being in a given class/status can be written with the aid of Bayesian theorem (Ebeling 2010):

$$P\left(\omega_k \mid \vec{X}\right) = \frac{P\left(\omega_k\right) P\left(\vec{X} \mid \omega_k\right)}{P\left(\vec{X}\right)} \qquad (1)$$

where

$$P\left(\vec{X}\right) = \sum_{k=1}^{c} P\left(\omega_k\right) P\left(\vec{X} \mid \omega_k\right) \qquad (2)$$

The probabilities of each class $P(\omega_k)$ are known as priors, $P(\vec{X} \mid \omega_k)$ are the conditional likelihoods, and $P(\vec{X})$ is the unconditional likelihood. The working status of the machine/system could be assessed by collecting data from various pseudo-discrete features x_i. The assessment process would be simple provided that a sufficient, if not abundant, amount of data is available. In practice, however, not every personnel running such complex machineries could make a judgement on the working status based on recorded data.

Take an air compressor as an example. For simplicity, we can define three possible classes for the compressor: normal operation, air leak, and overheating. (ω_1 , ω_2 and ω_3) To monitor these classes, we measure four pseudo-discrete features: air pressure, air flowrate, oil pressure and oil flowrate, i.e. $\vec{X} = (x_1, x_2, x_3, x_4)^T$.

Measurements from each pseudo-discrete feature are recorded. An operator may have to make a judgement based on the information presented in Table 1. By looking at the numbers presented in **Table 1**, it is exceedingly difficult for anyone to decide as of which class the compressor is working under. However, the decision making would be much easier if the person is presented with system information as in Table 2.

Table 1. Numerical Data Presentation

Air Pressure x_1	200 kPa
Air Flowrate x_2	1.5 kg/s
Oil Pressure x_3	50 kPa
Oil Temperature x_4	40 degrees Celsius

By looking at the numbers presented in **Table 1**, it is exceedingly difficult for anyone to decide as of which class the compressor is working under. However, the decision making would be much easier if the person is presented with system information as in **Table 2**.

Table 2. Pseudo-Discrete Presentation

Air Pressure x_1	Normal
Air Flowrate x_2	Normal
Oil Pressure x_3	High
Oil Temperature x_4	Critical

Table 2 presents a much clearer picture of the system's working status. Even a person not familiar with the compressor's normal working condition could tell that the compressor is likely overheating. Therefore, we aim to build a technical diagnostic system, using simple pseudo-discrete features that are easy to understand, with the aid of some expert knowledge and learning sample.

Expert Knowledge

To build a diagnostic system based on Table 2, each pseudo-discrete feature x_i is categorized further into pseudo-discretes $\mu_{j,i}$ (meaning the j^{th} pseudo-discrete of the i^{th} pseudo-discrete feature), describing the measurement in "levels" such as "too high", "high", "normal", "low" and "too low". The pseudo-discretes are divided using expert knowledge.

Using the same air compressor example, the pseudo-discrete feature oil temperature could be divided into four levels: low, normal, high and critical, with corresponding temperature ranges assigned as shown in Table 3 and Figure 1.

Table 3. Division of pseudo-discretes for pseudo-discrete feature Oil Temperature

Low	Below 30 °C
Normal	40-60 °C
High	70-90 °C
Critical	Above 100 °C

The expert knowledge provides the level ranges for each pseudo-discrete feature. Such levels are a lot easier to assess than an unexplained/uncategorized data.

Note that the conditional probability of each pseudo-discrete feature can then be represented with h_i typical non-overlapping intervals $\left[D_j, U_j\right]$, which according to (Tenekedjiev et al. 2006) are:

$$P\left(x_i \in \left[D_j, U_j\right] \mid \omega_k\right) = q_{j,i}^{k(x)} \qquad (3)$$

For example, the conditional probability of the air compressor being in third class (overheating), with the fourth pseudo-discrete feature (oil temperature) being in the second pseudo-discrete h_2 (normal), is:

$$P\left(x_4 \in \left[40°C, 60°C\right] \mid \omega_3\right) = q_{2,4}^{3}$$

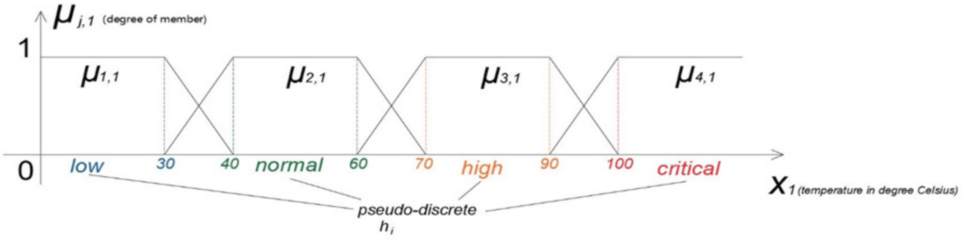

Figure 1. Fuzzy membership Function for Pseudo-discrete feature oil temperature with four pseudo-discretes

After consulting with an expert, the term $q_{j,i}^{k(e)}$ is gained as the reference probability of the i^{th} pseudo-discrete feature to be in the j^{th} pseudo-discrete while the system is working under class k. We can then build an expert knowledge vector for the i^{th} pseudo-discrete feature as:

$$\vec{E}_i^k = \left(q_{1,i}^{k(e)}, q_{2,i}^{k(e)}, q_{3,i}^{k(e)}, \ldots, q_{h_i,i}^{k(e)}\right) \qquad (4)$$

A confidence factor $L_d^{k(e)}$ is applied to the expert knowledge. Such factor is a measurement of the expert's accuracy or confidence when assigning probabilities to each pseudo-discrete with a given class. Applying the confidence factor to (4) gives the expert knowledge vector:

$$E_k = \left(\vec{E}_1^k - L_1^{k(e)}, \vec{E}_2^k - L_2^{k(e)}, \ldots, \vec{E}_d^k - L_d^{k(e)}\right) \qquad (5)$$

With each measured property now divided into pseudo-discretes for easier recognition, it is necessary to give the machine a learning sample to study and recognize the class of a system.

Learning Sample

For each pseudo-discrete feature, nk observations are given to the program for learning and recognition. A vector $\vec{X}_l^k$ consisting results of all pseudo-discrete features from the l^{th} observation in class k, is then given as:

$$\vec{X}_l^k = \left(X_{l,1}^k, X_{l,2}^k, X_{l,3}^k \ldots, X_{l,d}^k\right) \qquad (6)$$

For example, in the air compressor case, the vector $\vec{X}_{25}^3$ will be: $\vec{X}_{25}^3$=(101 kPa, 1.5 kg/s, 200 kPa, 88 ^{o}C) and represents readings from all four pseudo-discrete feature of the 25th observation, when the compressor is overheating. The air pressure reads 101 kPa, air flow rate reads 1.5 kg/s, oil pressure reads 200 kPa, and the oil pressure reads 88 C^{o}.

The term $q_{j,i}^{k(x)}$ from (3) is calculated using the frequentist definition of probability (Tenekedjiev et al. 2002)

$$q_{j,i}^{k(x)} = \frac{\sum_{l=1}^{nk} L_l^k \mu_{j,i}\left(X_{l,i}^k\right)}{\sum_{l=1}^{nk} L_l^k} \qquad (7)$$

with a goodness-of-fit factor L_l^k. The factor shows how well each observation truly represents the corresponding faults or class of a machine (Hald 2007).

Pseudo-discretes

There is a potential numerical problem with the conditional likelihood term $P\left(\vec{X} \mid \omega_k\right)$ in (1). This term is often a small value, and when it is smaller than the machine epsilon ε, it is treated as 0 in any machine language. To solve the stated numerical problem, $P\left(\omega_k \mid \vec{X}\right)$ is split into two terms by taking its logarithm:

$$ln\,P\left(\omega_k \mid \vec{X}\right) = A_k\left(\vec{X}\right) + B\left(\vec{X}\right) \qquad (8)$$

The $B\left(\vec{X}\right)$ part does not depend from the class *k*, whereas the part $A_k\left(\vec{X}\right)$ is different for each class. It is called the discriminant function for class *k*. That name originates from the fact that we can easily identify the class with the greatest posterior probability as the class with the greatest discriminant function (i.e. we can classify the observation $\vec{x}$ using the maximum posterior probability method based only the discriminant functions): $x \in \omega_k$ if $A_k\left(\vec{X}\right) \geq A_i\left(\vec{X}\right)$ for *i*=1,2,3,…, *c*

Using the discriminant functions allows us to avoid any numerical problems. Although we will never calculate the part $B\left(\vec{X}\right)$, it is trivial to derive an expression for the posterior probabilities depending only on the discriminant functions:

$$P\left(\omega_k \mid \vec{X}\right) = \frac{1}{\sum_{j=1}^{c} e^{A_j\left(\vec{X}\right) - A_k\left(\vec{X}\right)}} \qquad (9)$$

Taking the logarithm of $P\left(\omega_k \mid \vec{X}\right)$ in equation (1) we have:

$$ln\,P\left(\omega_k \mid \vec{X}\right) = ln\left(\frac{P\left(\omega_k\right)P\left(\vec{X} \mid \omega_k\right)}{P\left(\vec{X}\right)}\right) = ln\,P\left(\omega_k\right) + ln\,P\left(\vec{X} \mid \omega_k\right) - ln\,P\left(\vec{X}\right) \quad (10)$$

Independent of the class number *k*, it is then recognized that the term $-ln\,P\left(\vec{X}\right)$

is to be the term $B\left(\vec{X}\right)$, and the two terms $ln\,P\left(\omega_k\right)+ln\,P\left(\vec{X}\,|\,\omega_k\right)$ are to be the term $A\left(\vec{X}\right)$ in (8).

Assuming the measured pseudo-discrete features x_i are all independent from each other, then:

$$P\left(\vec{X}\,|\,\omega_k\right)=\prod_{i=1}^{d}P\left(x_i\,|\,\omega_k\right)=P\left(x_1\,|\,\omega_k\right)P\left(x_2\,|\,\omega_k\right)P\left(x_3\,|\,\omega_k\right)\ldots P\left(x_d\,|\,\omega_k\right) \quad (11)$$

The term $P\left(x_i\,|\,\omega_k\right)$ is of interest. If the measured feature X_i is assumed to be a pseudo-discrete feature with h_i pseudo-discretes, then it is possible to represent the theoretical justification of the fuzzy measurements, which leads us back to the setup of pseudo-discrete features and the introduction of expert knowledge section.

Pseudo-Bayesian Estimation

Combining the learning sample $q_{j,i}^{k(e)}$, and the expert knowledge $q_{j,i}^{k(e)}$, the final probability $q_{j,i}^{k}$ can then be estimated using the frequentist definition of probability with Pseudo-Bayesian Estimation, developed from (7):

$$q_{j,i}^{k}=\frac{\sum_{l=1}^{nk}L_l^k q_{j,i}^{k(x)}+50L_i^{k(e)}q_{j,i}^{k(e)}}{\sum_{l=1}^{nk}L_l^k+50L_i^{k(e)}} \quad (12)$$

The expert knowledge terms in (12) are applied with an accuracy factor that is subject to change. Here, 50 represents 1/50=2% accuracy factor (Skaggs 1989).

Epsilon Correction

It is possible that a certain observation would contain no data in certain pseudo-discretes, a.k.a. $q_{j,i}^{k}=0$. In this case, the zero probability is substituted with the machine epsilon ε, while the probabilities from other pseudo-discretes are multiplied by $1-\varepsilon$, so that the sum of probabilities form pseudo-discretes in the same observation still equal to one.

Using the compressor example again, in an observation from the 4th class "Overheating" (*k*=4), the 4th pseudo-discrete feature "Oil Temperature" (*i*=4) has 4 pseudo-discretes (low, normal, high, critical *j*=4). However, no data falls under the "critical" level. In this case, epsilon correction is performed as shown in Table 4.

Table 4. Epsilon Correction for $q_{j,4}^{4}$

Pseudo-Disrete	Original Observation	Epsilon-Corrected Observation
$q_{1,4}^{4}$	0.2	$0.2(1-\varepsilon)$
$q_{2,4}^{4}$	0.3	$0.3(1-\varepsilon)$
$q_{3,4}^{4}$	0.5	$0.5(1-\varepsilon)$
$q_{4,4}^{4}$	0	ε
Sum	1	1

Application on Marine Diesel Generator

With the established theoretical background, we apply the Bayesian classifier within a numerical example of a hypothetical MTU 8V396 marine diesel generator. The data for our numerical example is obtained from an expert. A total of 9 classes are established, monitored through 23 pseudo-discrete features, i.e., *k*=9, and *i*=23. The classes and features are listed in Table 5 and Table 6.

Table 5. List of Classes for the Marine Diesel Generator Example

ω_1	Metal Fatigue
ω_2	Lost of DC Voltage
ω_3	Insufficient Output Frequency
ω_4	Single Phase Voltage Drop
ω_5	Misalignment
ω_6	Faulty Knock in Bore
ω_7	Incorrect Air/Fuel Ratio
ω_8	Cooler Overheating
ω_9	Normal Operation

After consulting with an expert, the pseudo-discrete features are classified into 3 to 5 different pseudo-discretes, with ranges $\left[D_j, U_j\right]$ given to each pseudo-discrete, and expert knowledge $q_{j,i}^{k(e)}$ given to every pseudo-discrete of every pseudo-discrete feature under every class. A learning sample, containing 10 observations in each class are given to the Bayesian classifier for learning and recognition. The confidence factor applied to the expert knowledge is set at 100% for this analysis.

To demonstrate the parameter estimation methods and to test the performance of the classified the expert also provided 10 pseudo-observations to each class. Some of the observations are purposely put out of $\left[D_j, U_j\right]$ for some pseudo-discrete features to see if the classifier would recognize them as being in a different class.

After recognition, the Bayesian classifier produces an error matrix that can be summarized as follows:

ω_1: 8 observations recognized correctly;
2 observations recognized in class 9;
ω_2: 8 observations recognized correctly;
1 observation recognized in class 8;
1 observation recognized in class 9;
ω_3: 5 observations recognized correctly;
3 observations recognized in class 4;
2 observations recognized in class 9;
ω_4: All observations recognized correctly;
ω_5: 9 observations recognized correcly;
1 observation recognized in class 9;
ω_6: 8 observations recognized correctly;
1 observation recognized in class 5;
1 observation recognized in class 9;
ω_7: 9 observations recognized correctly;
1 observation recognized in class 9;
ω_8: All observations recognized correctly;
ω_9: 9 observations recognized correctly;
1 observation recognized in class 8.

Table 6. List of Pseudo-Discrete Features
for the Marine Diesel Generator Example

x_1	DC Voltage (V)
x_2	Oil Pressure (psi)
x_3	Oil Flowrate (L/min)
x_4	Oil Temperature (K)
x_5	Water Temperature (K)
x_6	Water Flowrate (L/min)
x_7	Boost Pressure (bar)
x_8	Boost Temperature 1 (K)
x_9	Boost Temperature 2 (K)
x_{10}	Speed (rpm)
x_{11}	Drive-end tri-axel Accelerometer x (mm/s)
x_{12}	Drive-end tri-axel Accelerometer y (mm/s)
x_{13}	Drive-end tri-axel Accelerometer z (mm/s)

x_{14}	Non-drive-end tri-axel Accelerometer x (mm/s)
x_{15}	Non-drive-end tri-axel Accelerometer y (mm/s)
x_{16}	Non-drive-end tri-axel Accelerometer z (mm/s)
x_{17}	Output Frequency (Hz)
x_{18}	Bank A Knock Censor (mm/s)
x_{19}	Bank B Knock Censor (mm/s)
x_{20}	U Single-Phase AC Voltage (V)
x_{21}	V Single-Phase AC Voltage (V)
x_{22}	W Single-Phase AC Voltage (V)
x_{23}	Fule Flowrate (L/hr)

Conclusion

The process of applying pseudo-discrete features to complex systems could drastically simplify the technical diagnostics process. Our Bayesian classifier was able to recognize the state of the hypothetical generator with few errors on the 120 pseudo-observations provided by the expert. If the given machine is working under a certain class, with sufficient and accurate/confident expert knowledge, the classifier is able to accurately recognized the pattern within the measured pseudo-discrete features.

The use of pseudo-discrete features improves the quality of education in marine engineering, where students need pattern classification in technical diagnostics of marine equipment. These features are easy to use and comprehend. The pattern classification process is more transparent in that way because students can track the diagnostics decisions to their knowledge on how the marine equipment operates.

Future tests of the system should include actual recorded data from running the generator under different conditions, and a few expert knowledge should be made unavailable. The classifier would have to learn from two different scenarios: only learning sample data available with no expert knowledge; only expert knowledge available with no learning sample. In these cases, there is the ability to recognized a pattern by combining known and unknown information and then predict the working class of generator/system using test run data.

REFERENCES

Duda, R., Hart, P., Stork, D., 2001. *Pattern classification*. Second Edition. Wiley.

Ebeling, C. E., 2010. *An Introduction to Reliability and Maintainability Engineering* (2nd ed.). Long Grove, Illinois, United States: Waveland Press Inc.

Hald, A., 2007. A history of parametric statistical inference form Bernoulli for Fisher, 1713 to 1935., In: *Sources and Studies in the History of Mathematics and Physical Sciences.* NY USA: Springer Verlag.

Koc, L., Mazzuchi, Th., Sarkani, Sh., 2012. A network intrusion detection system based on a Hidden Naïve Bayes multiclass classifier. *Expert Systems with Applications* **39**(18), 13492 – 13500.

Nikolova, N., Hirota, K., Kolev, K., Tenekedjiev, K., 2019. Technical diagnostic system in the maintenance of turbomachinery for ammonia synthesis in the process industries. *Journal of Loss Prevention in the Process Industries* **58**, 102 – 115.

Skaggs, G., Stevenson, J., 1989. A Comparison of Pseudo-Bayesian and Joint Maximum Likelihood Procedures for Estimating Item Parameters in the Three-Parameter IRT Model. *Applied Psychological Measurements*. **13**(4), 391 – 402.

Tenekedjiev, K., Dimitrakiev, D., Nikolova, N., 2002. Building frequentist distributions of continuous random variables. *Machine Mechanics*. 47, pp. 164 – 168.

Tenekedjiev, K., Kobashikawa, C., Nikolova, N., Hirota, K., 2006. Generic database for hybrid bayesian pattern recognition. *Journal of Advanced Computational Intelligence and Intelligent Informatics*. **10**(3), 419 – 431.

✉ **Guixin Fan**
https://orcid.org/0000-0002-0334-2271
Australian Maritime College
University of Tasmania
Launceston, TAS, Australia
E-mail: gfan@utas.edu.au

✉ **Natalia Nikolova**
ORCID iD: 0000-0001-6160-6282
1Australian Maritime College
University of Tasmania
Launceston, TAS, Australia

Nikola Vaptsarov Naval Academy
Varna, Bulgaria
E-mail: Natalia.Nikolova@utas.edu.au

✉ **Ty Smith**
https://orcid.org/0000-0001-6803-7967
Australian Maritime College
University of Tasmania
Launceston, TAS, Australia
E-mail: tasmith9@utas.edu.au
ty@nmis.com.au

✉ **Kiril Tenekedjiev**
ORCID iD: 0000-0003-3549-0671
Australian Maritime College
University of Tasmania
Launceston, TAS, Australia

Nikola Vaptsarov Naval Academy
Varna, Bulgaria
E-mail: Kiril.Tenekedjiev@utas.edu.au