Parameterisation of domain knowledge for rapid and iterative prototyping of knowledge-based systems

Andrew Young [a,*], Graeme West [a], Blair Brown [a], Bruce Stephen [a], Andrew Duncan [b], Craig Michie [a], Stephen D.J. McArthur [a]

[a] Institute for Energy and Environment, Department of Electronic and Electrical Engineering, University of Strathclyde, 204 George Street, Glasgow, G1 1XW, UK
[b] The Alan Turing Institute, British Library, 96 Euston Rd, London NW1 2DB, UK

ARTICLE INFO

ABSTRACT

In critical infrastructure applications, timely and consistent fault detection and diagnosis is an increasingly important operational process, especially in the energy sector where safety is of the utmost importance. To realise this, engineers have to manually analyse data acquired from several assets using predefined diagnostic processes, but this is a time-consuming process requiring significant amounts of specialist expert knowledge. Data-driven approaches to support fault detection and diagnosis, and other similar problems, can produce accurate results comparable to what the engineers can achieve in a fraction of the time. However, the majority of these data-driven techniques are black box techniques and lack explainability which is often necessary for explaining decisions about critical assets in the power generation industry. Knowledge-based systems, such as rule-based expert systems have been shown to provide not only accurate decisions but also the explanation and reasoning behind these decisions in some related applications. However, there is a significant time cost associated with the development of knowledge-based systems, and in particular with the knowledge elicitation process, where the domain expert's knowledge is formalised and is encoded into the system. This challenge is commonly referred to as the knowledge elicitation bottleneck.

In this paper, we present a novel approach to performing the knowledge elicitation using a set of symbolic primitives (rise, fall, fluctuate, and stable) to parameterise typical time-series condition monitoring data. The knowledge is represented by using a common language that can easily be communicated with (and from) the domain experts. This allows for the quick and accurate elicitation of the domain experts knowledge, but also the formalisation and implementation of the knowledge into a rapidly produced diagnostic system. Further to this, due to the parametrisation of the knowledge, it is possible to iteratively improve the knowledge base by updating these parameters based on new unseen data. This approach was applied to the Tennessee Eastman dataset, which is simulated data of a real-world industrial process. It was found that by using this approach it was possible to accurately and quickly capture the knowledge required to detect several faults within the case study dataset, but also provided fully explained reasons why each fault was detected by relating the explanations to the symbolic primitives previously defined.

1. Introduction

For many critical assets across the power generation sector, accurate and efficient fault detection and diagnosis are extremely important. This is due to both the overriding need to ensure safe continued operation of the asset and the costs associated with downtime of the plant, through reduced or zero power generation but also with potentially expensive and time-consuming repairs or replacements, and also the cost related to the time associated with diagnosing said fault.

Evaluation of faults is traditionally a time-consuming, manually intensive process and automation of this process provides the opportunity to introduce efficiency savings. To ensure the adoption of any automated approach similar or improved accuracy must be achieved over what can already be achieved manually by the engineers. Two main approaches to automating fault diagnosis are; data-driven approaches and knowledge-based approaches. With the increases in computing power available over the last two decades, it has been observed in Liu

et al. (2018) that there has been a huge swing to more data-driven approaches over the traditional knowledge-based approaches. However, for many critical assets in power generation there is often a requirement to provide supporting evidence when making decisions and before implementing related solutions, due to the costs associated with reduced power generation, or expensive repairs or replacements. Hence, the lack of explainability of many data-driven approaches is currently a problem that is being extensively researched (Chakraborty, Başağaoğlu, and Winterle (2021) and Moradi and Samwald (2021)), however, this is a problem that is yet to be solved. In addition to this, due to many of these types of assets having conservative design tolerances, to compensate for and avoid lost revenue during downtime, there is frequently a lack of failure case studies to train data-driven techniques and ensure their accuracy or effectiveness.

Knowledge-based techniques present an opportunity to address both the lack of: explainability; and, training data by capturing and codifying the domain expert's knowledge and expertise. Expert knowledge is a direct replacement for explainability as diagnostic results can be traced back to the knowledge from which it was derived. Training of models is not required as the model, in this case, is one of the expert's knowledge, which is assumed to contain years of experience and training. This knowledge in many cases has already been recorded through documented diagnostic procedures, however, this is not always complete enough to be formalised into the rules for a knowledge-based system. The main drawback of this type of approach is the time taken to elicit this required knowledge from the engineers to enable this approach. In Cullen and Bryman (1988) this has been coined the "knowledge elicitation bottleneck".

This paper presents a novel approach for tackling elements of the knowledge elicitation bottleneck for fault diagnosis problems and provides a case study of its use with a publicly available engineering data set. The next section of this paper provides more details of the various approaches to automated fault diagnosis, with an overview of the trade-offs between data-driven and knowledge-based approaches. This is important as most data driven approaches, by their nature lack the explainability inherent in knowledge based approaches have. Although, for knowledge-based approaches to become more viable the time taken to develop these systems needs to be addressed. Section three presents the various current approaches to perform knowledge elicitation for the development of knowledge-based systems. A new streamlined approach to performing knowledge elicitation using symbolic primitives is presented in section four, along with a step by step process for how to perform a knowledge elicitation session using this approach. This looks to address the issue of the time taken in the development of these systems, specifically in the knowledge elicitation phase. Section five demonstrates the proposed approach for knowledge elicitation and the development of a prototype expert system using a condition monitoring case study dataset. Conclusions and future work are presented in section six.

2. Background

2.1. Fault diagnosis

Fault diagnosis in many industrial applications involves engineers following a predefined diagnostic procedure. These procedures have typically developed over many years based on physical laws, industry standards and the engineers understanding of the specific asset under analysis and how the individual process variables, e.g. temperature, vibration, pressure, flow, current, voltage, etc., change under different fault conditions. Based on these factors, the engineers will look for features, or trends in the specific process variable to determine if a fault condition has been met, then by cross-referencing this with the current operating conditions of the plant will determine if any actions are required to be undertaken.

Fig. 1 shows an example of an industrial fault detection (or troubleshooting) procedure (Asturias & Gagen, 2007). This process involves splitting the troubleshooting process into two stages, first identifying the issue and secondly performing the actual repair. The seven steps that must be taken to identify the fault are; gather information, verify the issue, try quick fixes, use appropriate diagnostics, perform a split-half search, use additional resources to research the issue, and escalate the issue (if necessary). In this example diagnosis process, after gathering the relevant data the system has to be assessed by an expert, as referenced by "schedule service call", to verify the problem, failing any quick fixes resolving the issue, an extensive diagnostic procedure has to be undertaken involving; running diagnostics, performing a split-half search, and general research into the problem. Each of these steps can be a very time-consuming process for the experts, therefore, it is advantageous to automate as much of this process as possible.

The main parts in these types of processes that can be automated relate specifically to the assessment of gathered information or data, running diagnostics, assessing how the data changes over time, and how the relationship between various datastreams change all under different fault conditions.

The two approaches discussed in this paper to automate these steps are categorised into data-driven approaches and knowledge-based approaches.

2.2. Data-driven approaches

There are several data-driven approaches to automated fault diagnosis, however, the current state of the art is predominately Machine Learning oriented.

A convolutional neural network-based approach to fault diagnosis is proposed in Wen, Li, Gao, and Zhang (2018), the authors highlight the limitation of the approach being able to classify common fault conditions, i.e. those that have a large amount of labelled training data available. Therefore, for faults that have not been learned the technique would be unable to classify them. The authors of Guo, Lei, Xing, Yan, and Li (2019) propose a deep convolutional transfer learning network (DCTLN) method for fault diagnosis of unlabelled data. The results show that the DCTLN trained on labelled data from one machine could accurately classify unlabelled data acquired for a new machine. As before, this approach requires labelled data to exist for a machine similar to the machine currently being analysed. In Ding, Ma, Ma, Wang, and Lu (2019) the authors propose a generative adversarial network-based approach in an attempt to solve the "small sample size problem", by producing an augmented training dataset that was used to accurately classify the given faults. Also in Meng, Guo, Pan, Sun, and Liu (2019), the authors propose a new data augmentation by segmenting the sample data and randomly recombining the segments to artificially produce a larger amount of training data.

From the literature, for industrial fault diagnostics for a critical asset where there is extremely limited labelled faulty training data (if any at all) and supporting evidence is required when making diagnostic conclusions; in general, machine learning or data-driven approaches are unable to solve this type of problem. This is mainly due to the large amount of training data required to learn the faults or the lack of explainability due to the black-box nature of these techniques. In addition to this, for these types of industrial problems, there is a significant amount of domain expert knowledge available due to the current engineers having to perform this fault diagnosis process manually. The process of incorporating the domain expert knowledge into data-driven models is currently an active area of research but is far from a trivial problem (Deng, Ji, Rainey, Zhang, & Lu, 2020).
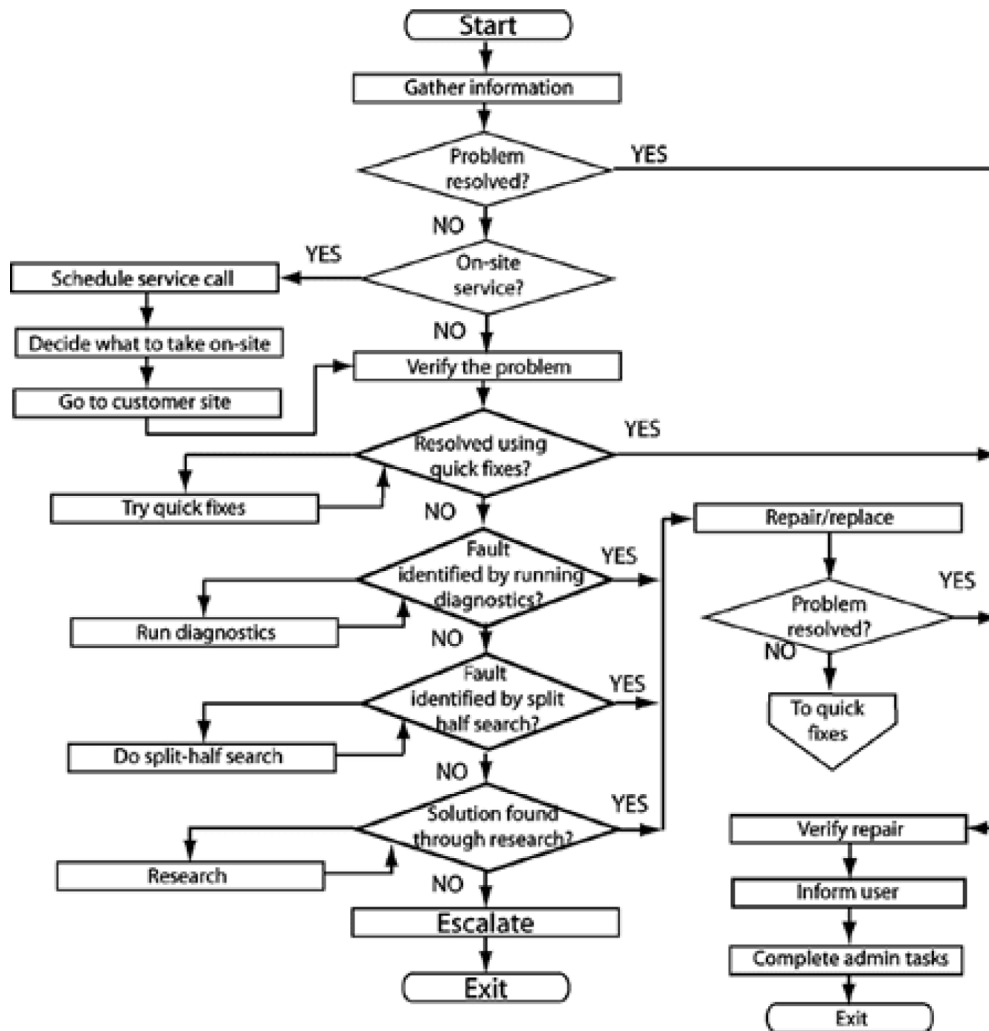
Fig. 1. General Troubleshooting Flowchart (Asturias & Gagen, 2007).

2.3. Knowledge-based approaches

An alternative class of methods to solve this problem involves knowledge-based approaches. Unlike data-driven approaches, knowledge-based approaches do not require a large amount of training data to be available, but instead, encode a significant amount of expert knowledge or human expertise into the system that can be used to perform fault diagnosis. As the domain experts knowledge is encoded into the system, they are not only able to produce diagnostic conclusions but are also able to explain the reasoning behind why a decision has been made.

In Qian, Li, Jiang, and Wen (2003) the authors present an expert system for fault diagnosis of chemical processes, using a pre-existing knowledge base. A Vibration-based expert system is presented in Yang, Lim, and Tan (2005) and uses a decision tree style approach to perform the analysis. This decision tree is generated from the cause-symptom matrix presented in Jackson et al. (1978). Due to the nature of how the rules are formalised and stored in a rule-based expert system, having low generality and low expandability (Gao, Cecati, & Ding, 2015) a task-based diagnostic expert system was proposed in Bo, Zhi-nong, and Zhong-qing (2012). This allows for the rules to be customised for specific assets and stored as more general rules, however, this also requires a well-structured logic to exist and applies only to similar assets.

As stated in Angeli (2010), Rafea, Hassen, and Hazman (2003), and Shadbolt and Smart (2015) the main hurdle to overcome in the

development of knowledge-based systems is the knowledge elicitation, or knowledge acquisition, bottleneck. Various research has been undertaken in an attempt to solve this problem (Chen & Rao, 2008; Wagner, Otto, & Chung, 2002; Xing, Huang, & Shi, 2003; Xiong, Litz, & Ressom, 2002), however, the majority of this research has focused on ways of extending the current knowledge elicitation techniques by the inclusion of more participants or additional data sources.

3. Knowledge elicitation

Knowledge elicitation is the process of attempting to elicit knowledge from domain experts using various types of methods and techniques. This information can be gathered from a variety of sources, e.g. technical manual, case studies, textbooks, etc., however, this is typically done by direct interaction with the domain expert. Generally, this is because the knowledge is often derived through practical experience. This implies that there is often the requirement for the expert to be involved in the knowledge elicitation process, representing the main challenge for the knowledge engineer. The person conducting the knowledge elicitation session is therefore required to determine an appropriate way for the experts to communicate their knowledge. The approach proposed in this paper aims to build on this premise by developing an approach that is both familiar to the domain expert, easy for them to communicate their understanding, and also easy to codify this knowledge into a knowledge base. Additionally, issues arise during knowledge elicitation, for example, due to the nature of the

knowledge being acquired by the expert over several years and hence becoming part of their routine, it can often, therefore, be difficult for the experts to formalise their knowledge. Also, for larger organisations, the knowledge can be distributed across a number of experts in various geographical locations or with different management oversight.

Various approaches can be used to perform this knowledge elicitation. The main techniques have been categorised into five main categories (Shadbolt & Smart, 2015); interviews, protocol analysis, diagramming, sorting and rating, and constrained processing which are shown below:

- Interviews
  - Structured
    * Fixed Probe (Shadbolt & Burton, 1990)
    * Focused Interviews (Hart, 1989)
    * Forward Scenario Simulation (Grover, 1983)
    * Critical Decision Method (Hoffman, 1998)
  - Semi-Structured
    * Knowledge Acquisition Grid (LaFrance, 1987)
    * Teach Back (Johnson & Johnson, 1987)
  - Unstructured (Weiss & Kulikowski, 1984)
- Protocol Analysis
  - Verbal
    * Online (Johnson, Zualkernan, & Garber, 1987)
    * Offline (Elstein, Shulman, & Sprafka, 1978)
    * Shadowing (Clarke, 1987)
    * Collegial Verbalization (Erlandsson & Jansson, 2007)
  - Behavioural (Ericsson & Simon, 1984)
- Diagramming
  - Laddered Grid (Corbridge, Rugg, Major, Shadbolt, & Burton, 1994)
  - Concept Mapping (Novak & Cañas, 2006)
  - Process Mapping (Milton, 2012)
- Sorting and Rating
  - Concept Sorting (Gammack, 1987)
  - Repertory Grid (Shaw & Gaines, 1987)
  - Pathfinder (Schvaneveldt et al., 1985)
- Constrained Processing
  - Limited-Information Task (Hoffman, 1987)
  - 20 Questions (Grover, 1983)

Each of these approaches has different positives and negatives, and a full discussion of this is beyond the scope of this paper. However, the main negative for all of these approaches is they are time-consuming for both the knowledge engineer and the domain expert. They also either require face to face meetings between both parties which takes time away from the day job of the domain experts, or requires access for the knowledge engineer to observe the domain expert performing the task, which for certain industries may not be possible.

4. Methodology

4.1. Background

In the literature, it is agreed that the knowledge elicitation bottleneck is a major problem to overcome before being able to rapidly develop and deploy industrial fault diagnostic expert systems. The novel approach proposed in this paper alters the typical expert system
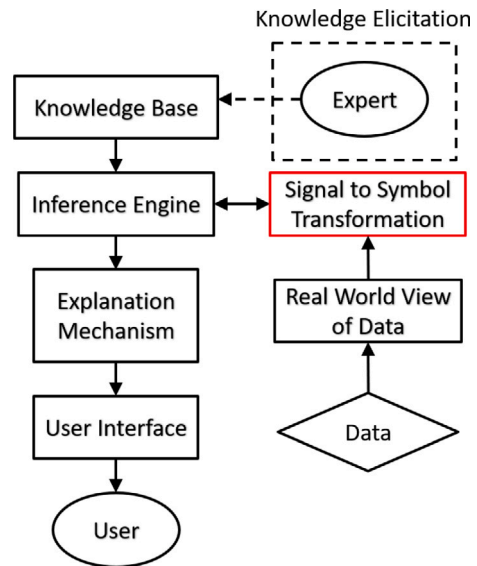

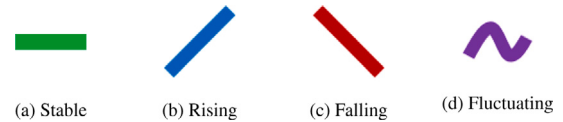**Fig. 2.** Updated expert system architecture, with signal to symbol transformation.

(a) Stable   (b) Rising   (c) Falling   (d) Fluctuating

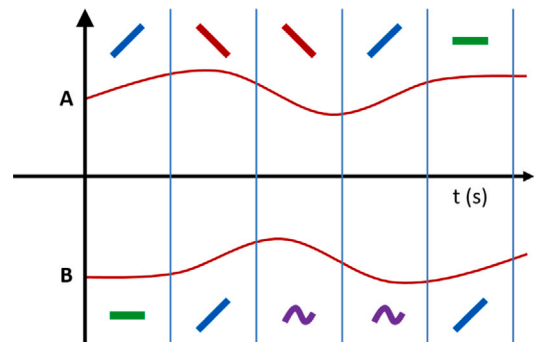**Fig. 3.** Four selected symbolic primitives.

Fig. 4. Example signal to symbol transformation of two time series data sources.

architecture by the inclusion of a signal to symbol transformation (SST) Costello, West, McArthur, and Campbell (2012) and Young, West, Brown, Stephen, and McArthur (2019) step. Fig. 2 shows the updated expert system architecture with the inclusion of an SST between the real world view of data and the inference engine.

SST is used to transform time series data gathered from the asset into a series of symbolic primitives that accurately represent the data. This is achieved by segmenting the data into discrete time intervals, then for each time segment a symbolic primitive is assigned, examples of, these primitives are intentionally rationalised (to minimise complexity) as: rising, falling, fluctuating or stable, see Fig. 3. However, the symbolic primitives can be customised for the specific application. An example of this applied to two-time series data sources is shown in Fig. 4.

The asset-specific knowledge that needs to be elicited from the domain experts is the subtle difference for each symbolic primitives. Each of the differences for each primitive has been assigned a parameter, this is shown in Fig. 5. Fig. 3(a) shows an example of how the x and y parameters are defined for a stable trend. Here the parameters define

Table 1
Example format of pressure datastream specific rules. (– is Stable, F is Fluctuating, ↑ is Rising and ↓ is Falling).

Name	D1	x	y	z	D2	x	y	z	D3	x	y	z	D4	x	y	z
Fault 1	–	1.5	2.0	N/A	↑	600	2.0	0.25	↓	600	1.5	0.5	F	600	10	N/A

(a) Stable

(b) Rising

(c) Falling

(d) Fluctuating

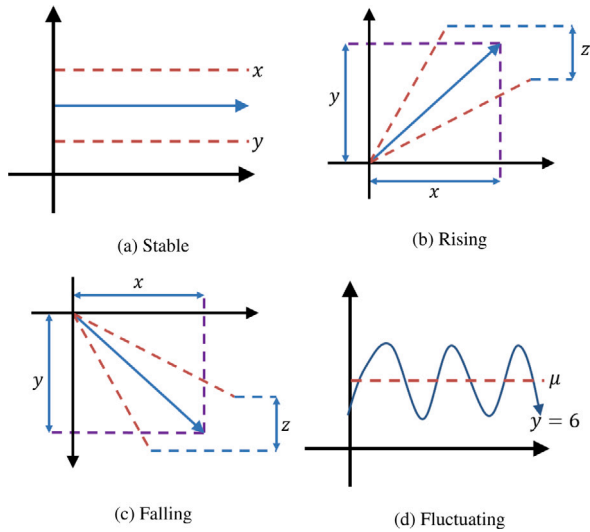**Fig. 5.** Definition of parameters for subtle difference in symbolic primitives. Where μ is the mean of the signal.

Algorithm 1: **Signal to symbol transformation. Where x, y and z are defined in Fig. 5**

if *50% of data ($< x*mean(data)$ or $> y*mean(data)$)* **then**
  | **Result:** Stable
else if *Number of mean crossings $> y$* **then**
  | **Result:** Fluctuating
else
  | Calculate average of first and last 10% of data for x period of time;
  | **if** *First $< y*Last \pm z*Last$* **then**
  | | **Result:** Rising
  | **else if** *First $> y*Last \pm z*Last$* **then**
  | | **Result:** Falling

the maximum deviation from the mean of the signal, both in terms of an increase and a decrease. Conversely, Fig. 3(d) gives an example of a fluctuating trend, for the data samples that fall outside of the stable regions, the number of fluctuations (mean crossings) is calculated if this is above the threshold defined by y the signal is considered fluctuating. For Fig. 3(b) and Fig. 3(c) the x, y and z parameters are defined the same except where y defined an increase for a rising trend and a decrease for a falling trend. The x parameter defines the time the trend occurs over and the z parameter is the expected deviation in the rise or fall of the trend.

By eliciting a combination of symbolic primitives for specific datastreams and eliciting for each of the x, y, and z parameters for each primitive it will be possible to construct a rule that can be directly interpreted by an expert system.

4.2. Knowledge elicitation session

In the previous subsection, an approach to alter the standard expert system architecture was described to include a SST step that allows for the parametrisation of the expert knowledge required to perform fault diagnosis. In addition to this, four example symbolic primitives were described and how each of these symbols were parametrised.

This section describes how framing the problem in this manner allows for more efficient domain knowledge capture from engineers in a knowledge elicitation session. In all of the knowledge elicitation approaches mentioned in Section 3, the knowledge engineer has to understand the process that the expert adopts when analysing their problem. The proposed approach simplifies the explanation of this process into several key questions; What datastreams are relevant to the given fault?, What trends are associated with the relevant datastreams?, What is the expected magnitude of these trends?, and How long will the fault take to manifest?

Below, we provide a step by step description of how to perform a knowledge elicitation session using the proposed approach.
0. Fault selection: The fault selection step, numbered step zero in the process as it occurs before the knowledge elicitation session. In this step, the domain experts must collate example case studies for each

fault they want the system to be able to detect. Ideally, this should be in a digital format (e.g. plots of process variables from a condition monitoring system) to allow for ease of sharing, and accurately defining of parameters in Step 4.
1. Datastream selection: The initial stage in the knowledge elicitation session, is datastream selection. Based on the case study data under investigation and supplied from the domain experts, all datastreams except the specific datastreams they would use to determine the fault are eliminated. In the case of the eliminated datastreams, these "Don't care" or "N/A" states are ignored and not referenced in the diagnostic process.
2. Time Interval selection: The second stage involves exploration of the case study data, to determine the approximate time frame/scale during which the fault will present or develop. That is the time from steady-state normal operation to a fault occurring, or being fully represented in the datastreams under analysis. Using the SST as described above, this time is used to segment the data. While this time interval does not need to be exact, and it is assumed that no two fault cases will be the same, a relatively accurate time interval is necessary to produce an accurate rule for the expert system.
3. Symbolic Primitive selection: For the next stage, each datastream is required to be assigned a symbolic primitive. Fig. 5 shows an example of four trends that can be used, however, additional primitives can be added to this list assuming appropriate parameters are defined for each new primitive.
4. Parameter Calculation: The final stage before a rule can be produced involves the setting of the individual parameters. Algorithm 1 shows pseudocode for how the symbolic primitive can be defined for individual time segments using the four previously defined primitives. As before this can be updated with the addition of new symbolic primitives.
5. Diagnostic Rule Induction: Implementation of the pseudocode shown in Algorithm 1, allows for the symbolic primitives to be calculated for any input datastream. A diagnostic rule can then be stored in the knowledge base for any number of datastreams, as shown in Table 1.

5. Case study: Tennessee eastman process simulation data

5.1. Background

The Tennessee Eastman process is a hypothetical set of interconnected industrial processes based on the physical behaviours of actual plant that was modelled computationally by Downs and Vogel (1993). The process contains five major units: the condenser, compressor,
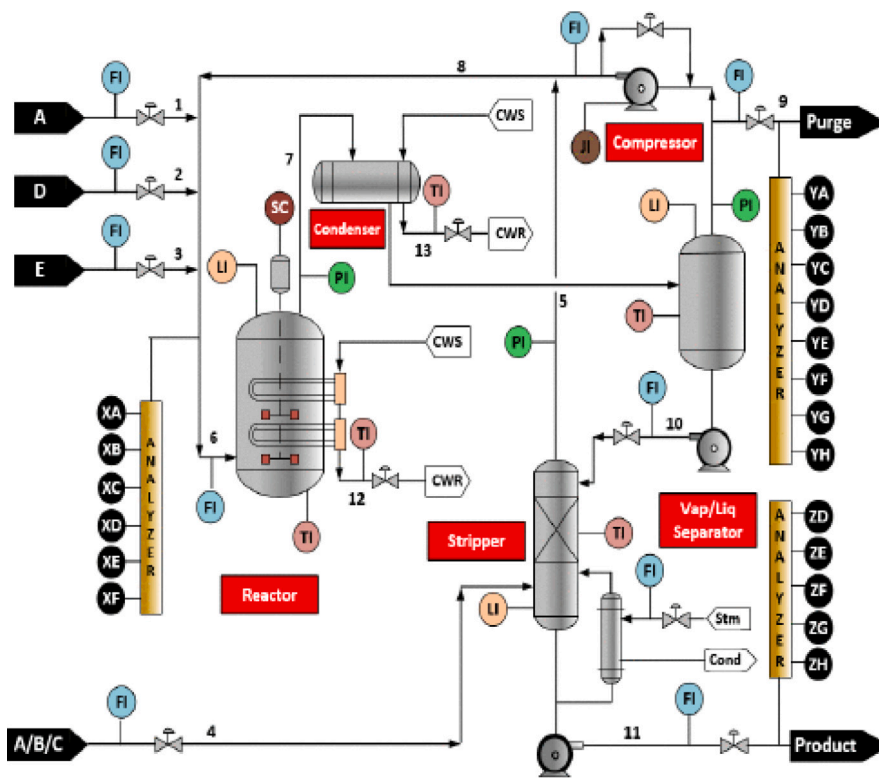
Fig. 6. Schematic diagram for the Tennessee Eastman Process (Ragab, El-koujok, Amazouz, & Yacout, 2017).

reactor, vapour/liquid separator and the product stripper, as shown in Fig. 6. This dataset has been used extensively in many areas including fault detection, plant-wide control, and statistical process monitoring. In Lomov, Lyubimov, Makarov, and Zhukov (2021) a temporal deep learning model was proposed and in Krishnannair and Aldrich (2017) the use of nonlinear singular spectrum analysis is proposed both for the purposes of fault detection. The dataset used in this section is the "TEP_Faulty_Training" portion of the Tennessee Eastman Process dataset (Averkiev, 2020). This contains 52 individual monitored process variables, detailed in Table 7, and 8. This portion of the dataset contains 500 simulations of 20 specific fault types each with 500 samples per fault. The sampling rate for the dataset was 3 min, therefore giving approximately 40 years worth of condition monitoring data with complete ground truth.

5.2. Knowledge capture

Using the process proposed in this paper the knowledge necessary to capture each of the five faults in the dataset was captured. For the application case study, as there were no domain experts involved in the process, all 52 datastreams were assessed visually to determine which datastreams were necessary to detect the given fault. To provide substantially different faults, fault numbers 1, 2, 5, 6, and 7 (as defined in the dataset) were selected for analysis. For Fault 1, this was assessed as xmeas_1, xmeas_4, xmeas_7, xmeas_13, xmeas_16, xmeas_19, xmeas_22, xmeas_34, xmv_3, xmv_4, and xmv_9, the individual time segments relating to fault 1 are shown in Fig. 7.

For each of the ten datastreams selected, three randomly selected examples of fault 1 were selected. The signal to symbol transformation was then performed. For the rising and falling trends the average rise or fall was taken as the y parameters, and the max variation was taken as the z parameter. Fluctuating trends are defined as the number of mean crossing from the segment of data, therefore, the y was defined as the max mean crossings out of the three examples. Finally, the stable trends were defined as the range where less than 50% of the data was

Table 2
Extracted knowledge for fault 1 formulated into rules. U is a rising trend, D is a falling trend, F is a fluctuating trend and S is stable.

Variable	Fault 1	Variable	Fault 1
xmeas_1	U, 500, 1.70	xmeas_19	U, 500, 0.05
xmeas_4	D, 500, 0.04	xmeas_34	U, 500, 0.01
xmeas_7	F, 500, 35	xmv_3	U, 500, 1.70
xmeas_13	F, 500, 35	xmv_4	D, 500, 0.03
xmeas_16	F, 500, 35	xmv_9	U, 500, 0.07

Table 3
Extracted knowledge for fault 2 formulated into rules. U is a rising trend, D is a falling trend, F is a fluctuating trend and S is stable.

Variable	Fault 2	Variable	Fault 2
xmeas_3	U, 500, 0.02	xmeas_34	D, 500, 0.60
xmeas_4	U, 500, 0.01	xmeas_35	U, 500, 0.01
xmeas_10	U, 500, 0.98	xmeas_36	U, 500, 0.02
xmeas_16	F, 500, 32	xmeas_39	D, 500, 0.25
xmeas_18	D, 500, 0.02	xmv_2	U, 500, 0.02
xmeas_19	D, 500, 0.40	xmv_6	U, 500, 0.97
xmeas_22	U, 500, 0.02	xmv_9	D, 500, 0.45
xmeas_28	D, 500, 0.62		

Table 4
Extracted knowledge for fault 5 formulated into rules. U is a rising trend, D is a falling trend, F is a fluctuating trend and S is stable.

Variable	Fault 5
xmeas_22	S, 500, 0.01
xmv_11	U, 500, 0.07

outside of the boundary. The extracted knowledge for Fault 1 is shown in Table 2. As there are 500 samples per fault in the dataset the time interval for each piece of knowledge was set to 500. This process was then repeated for the remaining 4 faults, and the extracted knowledge is shown in Tables 2 to 6.
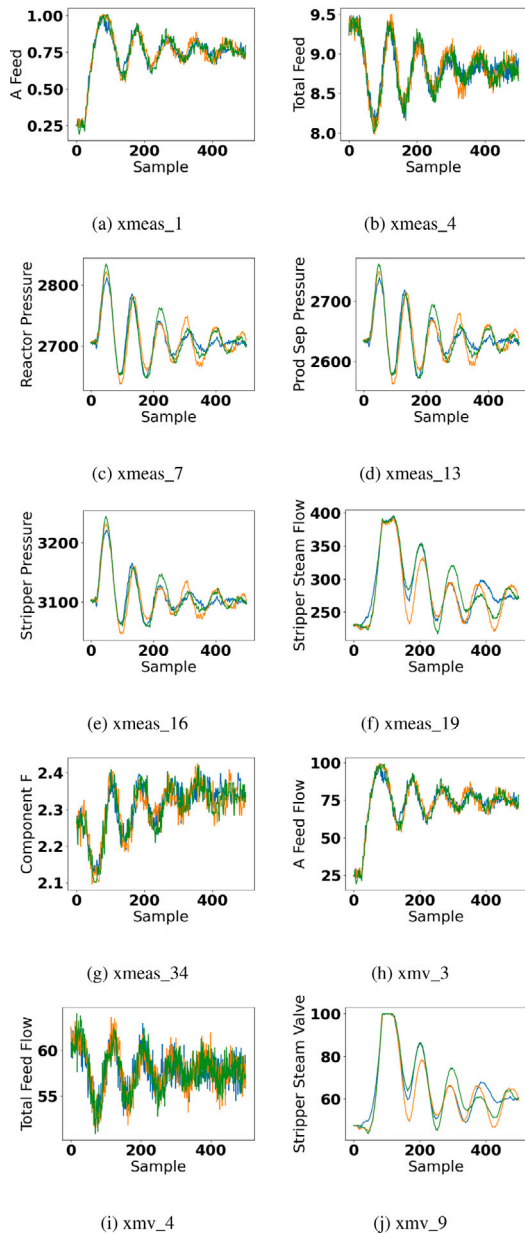
(a) xmeas_1

(b) xmeas_4

(c) xmeas_7

(d) xmeas_13

(e) xmeas_16

(f) xmeas_19

(g) xmeas_34

(h) xmv_3

(i) xmv_4

(j) xmv_9

Fig. 7. Captures of knowledge for fault 1.

Table 5
Extracted knowledge for fault 6 formulated into rules. U is a rising trend, D is a falling trend, F is a fluctuating trend and S is stable.

Variable	Fault 6	Variable	Fault 6
xmeas_3	D, 500, 0.03	xmeas_28	D, 500, 0.08
xmeas_7	U, 500, 0.10	xmeas_29	D, 500, 0.49
xmeas_10	D, 500, 0.34	xmeas_30	D, 500, 0.01
xmeas_11	D, 500, 0.08	xmeas_31	U, 500, 0.56
xmeas_13	U, 500, 0.10	xmeas_34	D, 500, 0.12
xmeas_16	U, 500, 0.10	xmeas_35	D, 500, 0.35
xmeas_18	U, 500, 0.02	xmeas_36	D, 500, 0.42
xmeas_19	U, 500, 0.96	xmeas_38	U, 500, 0.19
xmeas_20	D, 500, 0.09	xmv_2	D, 500, 0.01
xmeas_22	D, 500, 0.04	xmv_3	U, 500, 2.11
xmeas_23	D, 500, 0.28	xmv_4	U, 500, 0.21
xmeas_24	D, 500, 0.01	xmv_6	D, 500, 4.68
xmeas_25	U, 500, 0.32	xmv_9	U, 500, 1.06

The final step was to take all the data gathered in the above tables and implement each of them into a separate rule. Any datastreams not

Table 6
Extracted knowledge for fault 7 formulated into rules. U is a rising trend, D is a falling trend, F is a fluctuating trend and S is stable.

Variable	Fault 7
xmeas_22	S, 500, 0.01
xmv_4	U, 500, 0.19

mentioned in the tables were set as "don't care" states. This produced a knowledge-based expert system that contained 5 rules that were able to accurately detect Faults 1, 2, 5, 6 and 7 from the Tennessee Eastman Process dataset.

5.3. Results

Using this approach it was possible to produce a system that can accurately detect specific types of faults in the condition monitoring dataset. Out of the 10,000-time segments analysed only 8 (0.08%) false positives were produced, these were 6 from fault class 1 and 2 from fault class 2. Slightly more false positives were produced, 55 (0.55%) fault class 5 and 7 faults class 7 were detected as no-fault. However, as this system was being designed for fault diagnostics, more false positives and fewer false negatives was preferred. There was also 2 (0.02%) instances of fault class 1 that was detected as fault class 5, however, upon further analysis, it was visually impossible to differentiate these two instances for the general fault 5 class. Fig. 8 presents a confusion matrix for the full system.

6. Comparison with state-of-the-art

The methods highlighted in Section 3 are general approaches that can be applied to the capture of a wide range of knowledge from a broad range of domains. In this paper, we focus on the capture of a specific class of knowledge, that is associated with the interpretation of time series data. By constraining the type of knowledge being elicited allows additional structure to be placed around the representation of the captured knowledge, which in turn offers the opportunity to streamline the process.

Comparing approaches for the capture of human expertise is challenging, not least in that no two subject matter experts will respond to the interview process in the same manner, and attempting to elicit the same knowledge from the same expert, using two different methods in consecutive sessions is going to favour the latter method as the expert will be re-running through the same material. To attempt to address this issue a study was performed which compared the capture of a set of condition assessment rules derived for the health of an industrial asset through the examination of multiple streams of time series data. A subset of the rules pertaining to a specific fault mode were captured using a structured interview process, while a second subset of rules for a different fault mode was captured using the proposed methodology. In the former, 2 three hour knowledge elicitation sessions were held with the domain experts and it was possible to understand the problem, capture the relevant datastreams, and capture the relevant rules, however, within these two sessions it was not possible to parametrise the rules. For the same problem, in 1 two hour knowledge elicitation session it was possible to capture the relevant datastreams, capture the relevant rules, and parametrise the faults detected in four different case study datasets. While the specifics of the rules and parameters captured in each of these approaches are different, it was felt that they were suitably similar to permit valid comparisons in terms of the time required to be drawn. In conclusion, comparing the proposed technique with the standard structured interview, showed in this example a twofold saving in time.
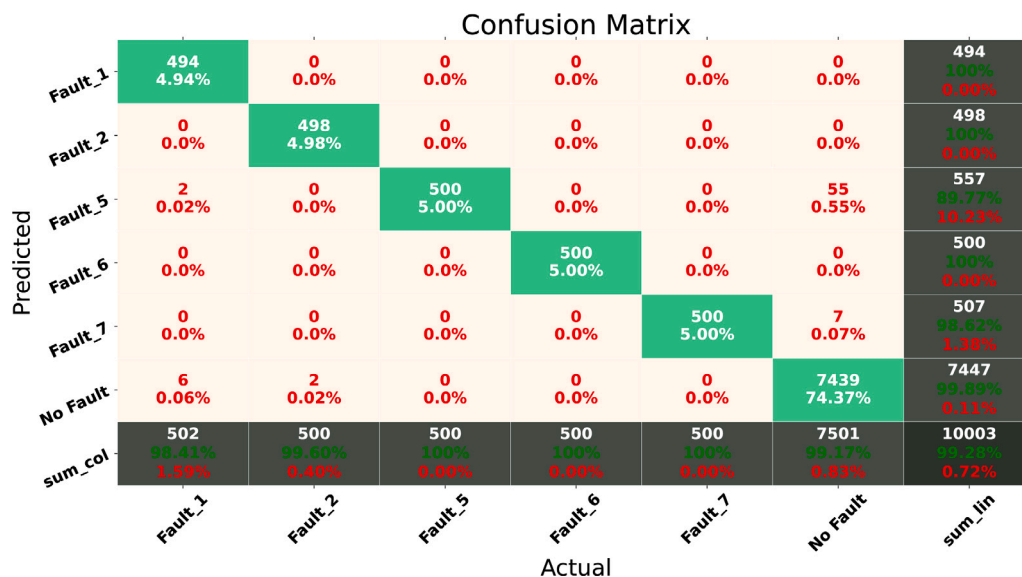
Fig. 8. Confusion matrix.

Table 7
Output variables of the Tennessee Eastman process.

Variable	Description
Xmeas_1	A feed (stream 1)
Xmeas_2	D feed (stream 2)
Xmeas_3	E feed (stream 3)
Xmeas_4	Total feed (stream 4)
Xmeas_5	Recycle flow (stream 8)
Xmeas_6	Reactor feed rate (stream 6)
Xmeas_7	Reactor pressure
Xmeas_8	Reactor level
Xmeas_9	Reactor temp
Xmeas_10	Purge rate (stream 9)
Xmeas_11	Separator temp
Xmeas_12	Separator level
Xmeas_13	Separator pressure
Xmeas_14	Separator underflow (stream 10)
Xmeas_15	Stripper level
Xmeas_16	Stripper pressure
Xmeas_17	Stripper underflow (stream 11)
Xmeas_18	Stripper temperature
Xmeas_19	Stripper steam flow
Xmeas_20	Compressor work
Xmeas_21	Reactor cooling water outlet temp
Xmeas_22	Condenser cooling water outlet temp
Xmeas_23 to	Composition of A to F
Xmeas_28	(stream 6)
Xmeas_29 to	Composition of A to H
Xmeas_36	(stream 9)
Xmeas_37 to	Composition of D to H
Xmeas_41	(stream 11)

Table 8
Input variables of the Tennessee Eastman process.

Variable	Description
Xmv_1	D feed flow (stream 2)
Xmv_2	E feed flow (stream 3)
Xmv_3	A feed flow (stream 1)
Xmv_4	Total feed flow (stream 4)
Xmv_5	Compressor recycle valve
Xmv_6	Purge valve (stream 9)
Xmv_7	Separator pot liquid flow (stream 10)
Xmv_8	Stripper liquid product flow
Xmv_9	Stripper steam valve
Xmv_10	Reactor cooling water flow
Xmv_11	Condenser cooling water flow
Xmv_12	Agitator speed

7. Conclusion and future work

This paper has introduced a new approach to tackle the knowledge elicitation process for a sub-set of problems, namely the codification and capture of knowledge involved in the interpretation of time series data. This type of problem is seen across many engineering domains and is particularly prevalent in the development of expert systems for supporting the health assessment of industrial assets. While knowledge-driven approaches, such as expert systems were prominent in the 1980's and 1990's advances in neural networks, and in particular the advent of deep neural networks have resulted in significant advances in many AI applications. However, for many engineering applications such as health monitoring and assessment of industrial assets, there is still the need for the explainability that is currently unavailable

from purely data drive black-box approaches. As a result, the capture and codification of specialist knowledge relating to the interpretation of condition monitoring data is of importance, and the issue of the knowledge elicitation bottleneck still exists.

This proposed approach has been shown to offer benefits in terms of time-saving. Historically, the main drawback to the development of these types of systems was the time involved in the capturing of domain expert knowledge from already time-poor engineers. The reduction in time, specifically in the knowledge elicitation phase, from a business perspective and the development of these types of systems places much less of a time burden on the already busy engineers. Future work will look at further reducing the time associated with capturing the domain expert knowledge. This could be achieved through the development of a learning algorithm that can process historical data and automatically propose new rules based upon repeating patterns that have been autonomously identified in historic data. This system could be implemented in a human-in-the-loop fashion to allow for the engineers to have the final say on what proposed rules are added to the knowledge base.

This new approach to performing knowledge elicitation for knowledge-based systems allows for the parameterisation of domain expert knowledge and the rapid prototyping of an expert system. This is important as there is an inherent brittleness in traditional expert systems due to the formalisation of the knowledge, and the ease in which the knowledge can be updated. It was demonstrated using a case study that it was possible to produce a knowledge-based system that can detect specific faults quickly and accurately. Also using this

approach to formulate the knowledge in a symbolic manner results in the explanation of the results in an accessible way that does not require any specialist training. Due to the nature of how an expert system is designed using this approach, updating specific parameters allows for the iterative improvement of the knowledge base and hence the system without the need to redesign the entire system. Future work will focus on the development of an incremental learning algorithm to propose improvements to existing domain knowledge based on user-defined false positives and negatives.

This paper represents a step towards reducing the knowledge elicitation bottleneck, but there are limitations to the proposed approach. It currently focuses only on the interpretation of time series data and anticipates the captured knowledge will reason about trends in data. Furthermore, temporal relationships between multiple data streams are relatively common in condition monitoring data but have not been fully addressed here. For example, seeing an increase in one parameter, flow rate, for example, may result in a corresponding, but delayed increase in value in another parameter, say temperature. Secondly, the elicitation process is still predominantly a human to human process, which is then transferred to the expert system. Significant research is still required to enable the domain expert to directly interface with the diagnostic system in a two-way manner, so that new knowledge can be imparted into the system directly by the expert as new, previously unseen faults arise. This data imbalance, where there are often large volumes of normal operating data, and very few instances of failure data is another common challenge, and capture and codification of this diagnostic knowledge in a manner that they can be shared across multiple instances of expert systems remains an open problem. Finally, the other significant area of research in this field will be the integration of the data-driven with knowledge-driven approaches, particularly if large volumes of operating data can be leveraged to provide increased confidence or evidence in the human-derived parametrised values.

CRediT authorship contribution statement

Andrew Young: Conceptualization, Methodology, Software, Validation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Graeme West:** Conceptualization, Methodology, Supervision, Writing – review & editing, Funding acquisition. **Blair Brown:** Conceptualization, Methodology, Writing – review & editing, Project administration. **Bruce Stephen:** Conceptualization, Methodology, Writing – review & editing. **Andrew Duncan:** Conceptualization, Writing – review & editing, Funding acquisition. **Craig Michie:** Conceptualization, Writing – review & editing, Funding acquisition. **Stephen D.J. McArthur:** Conceptualization, Writing – review & editing, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Appendix

See Tables 7 and 8.

References

Angeli, C. (2010). Diagnostic expert systems: From expert's knowledge to real-time systems. *Advanced Knowledge Based Systems: Model, Applications & Research, 1*.

Asturias, M., & Gagen, M. (2007). *Apple training series: desktop and portable systems*. Pearson Education.

Averkiev, S. (2020). Tennessee eastman process simulation dataset. URL: https://www.kaggle.com/averkij/tennessee-eastman-process-simulation-dataset/metadata.

Bo, M., Zhi-nong, J., & Zhong-qing, W. (2012). Development of the task-based expert system for machine fault diagnosis. *Journal of Physics: Conference Series, 364*, Article 012043. http://dx.doi.org/10.1088/1742-6596/364/1/012043.

Chakraborty, D., Başağaoğlu, H., & Winterle, J. (2021). Interpretable vs. noninterpretable machine learning models for data-driven hydro-climatological process modeling. *Expert Systems with Applications, 170*, Article 114498. http://dx.doi.org/10.1016/j.eswa.2020.114498.

Chen, C.-H., & Rao, Z. (2008). MRM: A Matrix representation and mapping approach for knowledge acquisition. *Knowledge-Based on Systems, 21*(4), 284–293. http://dx.doi.org/10.1016/j.knosys.2007.07.002.

Clarke, B. (1987). Knowledge acquisition for real-time knowledge-based systems. In *Proceedings of the first European workshop on knowledge acquisition for knowledge based systems* (pp. 2–3).

Corbridge, C., Rugg, G., Major, N., Shadbolt, N., & Burton, A. (1994). Laddering: technique and tool use in knowledge acquisition. *Knowledge Acquisition, 6*(3), 315–341.

Costello, J. J. A., West, G. M., McArthur, S. D. J., & Campbell, G. (2012). Self-tuning routine alarm analysis of vibration signals in steam turbine generators. *IEEE Transactions on Reliability, 61*(3), 731–740. http://dx.doi.org/10.1109/TR.2012.2209257.

Cullen, J., & Bryman, A. (1988). The knowledge acquisition bottleneck: time for reassessment? *Expert Systems, 5*(3), 216–225.

Deng, C., Ji, X., Rainey, C., Zhang, J., & Lu, W. (2020). Integrating machine learning with human knowledge. *IScience, 23*(11), Article 101656. http://dx.doi.org/10.1016/j.isci.2020.101656, URL: https://www.sciencedirect.com/science/article/pii/S2589004220308488.

Ding, Y., Ma, L., Ma, J., Wang, C., & Lu, C. (2019). A generative adversarial network-based intelligent fault diagnosis method for rotating machinery under small sample size conditions. *IEEE Access, 7*, 149736–149749.

Downs, J., & Vogel, E. (1993). A plant-wide industrial process control problem. *Computers & Chemical Engineering, 17*(3), 245–255. http://dx.doi.org/10.1016/0098-1354(93)80018-I, Industrial challenge problems in process control.

Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1978). *Medical problem solving: an analysis of clinical reasoning* (p. 346). Harvard University Press.

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: verbal reports as data*. the MIT Press.

Erlandsson, M., & Jansson, A. (2007). Collegial verbalisation–a case study on a new method on information acquisition. *Behaviour & Information Technology, 26*(6), 535–543.

Gammack, J. G. (1987). Different techniques and different aspects on declarative knowledge. In *Knowledge acquisition for expert systems* (pp. 137–163). Springer.

Gao, Z., Cecati, C., & Ding, S. X. (2015). A survey of fault diagnosis and fault-tolerant techniques—Part II: Fault diagnosis with knowledge-based and hybrid/active approaches. *IEEE Transactions on Industrial Electronics, 62*(6), 3768–3774.

Grover, M. D. (1983). A pragmatic knowledge acquisition methodology. In *IJCAI'83, Proceedings of the eighth international joint conference on artificial intelligence - Volume 1* (pp. 436–438). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc..

Guo, L., Lei, Y., Xing, S., Yan, T., & Li, N. (2019). Deep convolutional transfer learning network: A new method for intelligent fault diagnosis of machines with unlabeled data. *IEEE Transactions on Industrial Electronics, 66*(9), 7316–7325.

Hart, A. (1989). *New technology modular series, Knowledge acquisition for expert systems*. Kogan Page.

Hoffman, R. R. (1987). The problem of extracting the knowledge of experts from the perspective of experimental psychology. *AI Magazine, 8*(2), 53.

Hoffman, R. R. (1998). How can expertise be defined? Implications of research from cognitive psychology. In *Exploring expertise: issues and perspectives* (pp. 81–100). London: Palgrave Macmillan UK, http://dx.doi.org/10.1007/978-1-349-13693-3_4.

Jackson, C., et al. (1978). *The practical vibration primer*. Texas, USA: Gulf Publishing Company.

Johnson, L., & Johnson, N. E. (1987). Knowledge elicitation involving teachback interviewing. In *Knowledge acquisition for expert systems: a practical handbook* (pp. 91–108). Boston, MA: Springer US, http://dx.doi.org/10.1007/978-1-4613-1823-1_5.

Johnson, P., Zualkernan, I., & Garber, S. (1987). Specification of expertise. *International Journal of Human Computer Studies, 26*(2), 161–181. http://dx.doi.org/10.1016/S0020-7373(87)80089-5.

Krishnannair, S., & Aldrich, C. (2017). Fault detection in the Tennessee Eastman benchmark process with nonlinear singular spectrum analysis. *IFAC-PapersOnLine, 50*(1), 8005–8010. http://dx.doi.org/10.1016/j.ifacol.2017.08.1223, 20th IFAC World Congress.

LaFrance, M. (1987). The Knowledge Acquisition Grid: a method for training knowledge engineers. *International Journal of Man-Machine Studies, 26*(2), 245–255. http://dx.doi.org/10.1016/S0020-7373(87)80094-9.

Liu, J., Kong, X., Xia, F., Bai, X., Wang, L., Qing, Q., et al. (2018). Artificial intelligence in the 21st century. *IEEE Access*, *PP*, 1. http://dx.doi.org/10.1109/ACCESS.2018. 2819688.

Lomov, I., Lyubimov, M., Makarov, I., & Zhukov, L. E. (2021). Fault detection in Tennessee Eastman process with temporal deep learning models. *Journal of Industrial Information Integration*, *23*, Article 100216. http://dx.doi.org/10.1016/j. jii.2021.100216.

Meng, Z., Guo, X., Pan, Z., Sun, D., & Liu, S. (2019). Data segmentation and augmentation methods based on raw data using deep neural networks approach for rotating machinery fault diagnosis. *IEEE Access*, *7*, 79510–79522.

Milton, N. (2012). 11 Acquiring knowledge from subject matter experts. *Knowledge Service Engineering Handbook*, 253.

Moradi, M., & Samwald, M. (2021). Post-hoc explanation of black-box classifiers using confident itemsets. *Expert Systems with Applications*, *165*, Article 113941. http://dx.doi.org/10.1016/j.eswa.2020.113941.

Novak, J. D., & Cañas, A. J. (2006). The theory underlying concept maps and how to construct them. *Florida Institute for Human and Machine Cognition*, *1*(1), 1–31.

Qian, Y., Li, X., Jiang, Y., & Wen, Y. (2003). An expert system for real-time fault diagnosis of complex chemical processes. *Expert Systems with Applications*, *24*(4), 425–432. http://dx.doi.org/10.1016/S0957-4174(02)00190-2.

Rafea, A., Hassen, H., & Hazman, M. (2003). Automatic knowledge acquisition tool for irrigation and fertilization expert systems. *Expert Systems with Applications*, *24*(1), 49–57. http://dx.doi.org/10.1016/S0957-4174(02)00082-9.

Ragab, A., El-koujok, M., Amazouz, M., & Yacout, S. (2017). Fault detection and diagnosis in the Tennessee Eastman Process using interpretable knowledge discovery. In *2017 annual reliability and maintainability symposium (RAMS)* (pp. 1–7). http://dx.doi.org/10.1109/RAM.2017.7889650.

Schvaneveldt, R. W., Durso, F. T., Goldsmith, T. E., Breen, T. J., Cooke, N. M., Tucker, R. G., et al. (1985). Measuring the structure of expertise. *International Journal of Man-Machine Studies*, *23*(6), 699–728.

Shadbolt, N. R., & Burton, A. M. (1990). Experiments in knowledge elicitation. In K. MacGraw, & C. Westphal (Eds.), *Readings in knowledge acqusition: current practices and trends*. URL: https://eprints.soton.ac.uk/252364/.

Shadbolt, N., & Smart, P. (2015). Knowledge elicitation: methods, tools and techniques. (pp. 163–200).

Shaw, M. L., & Gaines, B. R. (1987). An interactive knowledge-elicitation technique using personal construct technology. In *Knowledge acquisition for expert systems* (pp. 109–136). Springer.

Wagner, W., Otto, J., & Chung, Q. (2002). Knowledge acquisition for expert systems in accounting and financial problem domains. *Knowledge-Based Systems*, *15*(8), 439–447. http://dx.doi.org/10.1016/S0950-7051(02)00026-6.

Weiss, S., & Kulikowski, C. (1984). *A practical guide to designing expert systems* (pp. 329–330). Rowman and Allanheld.

Wen, L., Li, X., Gao, L., & Zhang, Y. (2018). A new convolutional neural network-based data-driven fault diagnosis method. *IEEE Transactions on Industrial Electronics*, *65*(7), 5990–5998.

Xing, H., Huang, S., & Shi, J. (2003). Rapid development of knowledge-based systems via integrated knowledge acquisition. *AI EDAM*, *17*, 221–234.

Xiong, N., Litz, L., & Ressom, H. (2002). Learning premises of fuzzy rules for knowledge acquisition in classification problems. *Knowledge and Information Systems*, *4*, 96–111. http://dx.doi.org/10.1007/s10115-002-8195-4.

Yang, B.-S., Lim, D.-S., & Tan, A. C. C. (2005). VIBEX: An expert system for vibration fault diagnosis of rotating machinery using decision tree and decision table. *Expert Systems with Applications*, *28*(4), 735–742. http://dx.doi.org/10.1016/j.eswa.2004. 12.030.

Young, A., West, G., Brown, B., Stephen, B., & McArthur, S. (2019). Improved explicability for pump diagnostics in nuclear power plants. In *2019 ANS winter meeting and nuclear technology expo* (pp. 416–419).