Environment Systems and Decisions (2022) 42:177–188
https://doi.org/10.1007/s10669-021-09841-z

A novelty-based multi-objective evolutionary algorithm for identifying functional dependencies in complex technical infrastructures from alarm data

Federico Antonello[1] · Piero Baraldi[1] · Enrico Zio[1,2] · Luigi Serio[3]

Accepted: 27 December 2021 / Published online: 15 January 2022

Abstract
In this work, a Multi-Objective Evolutionary Algorithm (MOEA) is developed to identify Functional Dependencies (FDEPs) in Complex Technical Infrastructures (CTIs) from alarm data. The objectives of the search are the maximization of a measure of novelty, which drives the exploration of the solution space avoiding to get trapped in local optima, and of a measure of dependency among alarms, which drives the uncovering of functional dependencies. The main contribution of the work is the direct identification of patterns of dependent alarms; this avoids going through the preliminary step of mining association rules, as typically done by state-of-the-art methods which, however, fail to identify rare functional dependencies due to the need of setting a balanced minimum occurrence threshold. The proposed framework for FDEPs identification is applied to a synthetic alarm database generated by a simulated CTI model and to a real large-scale database of alarms collected at the CTI of CERN (European Organization for Nuclear Research). The obtained results show that the framework enables the thorough exploration of the solution space and captures also rare functional dependencies.

Keywords Complex technical infrastructures · Functional dependencies · Alarm data · Abnormal behaviours · Multi-objective evolutionary algorithm · Genetic algorithm · Novelty search · Particle accelerator

Abbreviations

CERN	European Organization for Nuclear Research
LHC	Large Hadron Collider
ARM	Association Rule Mining
EA	Evolutionary Algorithm
GA	Genetic Algorithm
MOEA	Multi-Objective Evolutionary Algorithm
CTI	Complex Technical Infrastructure
FDEPs	Functional Dependencies

List of symbols

N_c	Number of CTI components
c_j	Generic j-th component
a_i^{kj}	Alarm associated to the k-th malfunction of the j-th component
M_i^{al}	Number of types of alarm messages triggered by the j-th component
M^{al}	Total number of types of alarm messages
$A = \{a_i^{kj}\}$	Set of all possible alarm types
N^{al}	Total number of alarm messages collected in the database
$[t_0,t_f]$	Time domain during which the N^{al} alarm messages of the database have been collected
Z	Number of time intervals in which the time domain $[t_0,t_f]$ is subdivided
Δt	Time interval length
$a_j^k(z)$	Boolean variable associated to the occurrence of the alarm
a_j^{kj}	In the z-th time interval
$\vec{c}_j(z)$	Vector of size M_j^{al} indicating the state of the j-th component in the z-th time interval
$\vec{\mathcal{T}}(z)$	Vector of size M^{al} indicating the state of the CTI in the z-th time interval
T	Matrix of size $[Z \times M^{al}]$ representing the evolution of the CTI state in the time domain $[t_0,t_f]$
X	Pattern of alarms

✉ Piero Baraldi
piero.baraldi@polimi.it

[1] Energy Department, Politecnico di Milano, Via Lambruschini 4, 20156 Milan, Italy

[2] MINES ParisTech, PSL Research University, CRC, 75006 Sophia Antipolis, France

[3] Technology Department, CERN, 1211 Geneva 23, Switzerland

$n(X)$	Number of time intervals in which at least all the alarms of X occur
$S(X)$	Support of X
$P(X)$	Probability of occurrence of X
X^{fp}	Frequent pattern of alarms
$s\%$	Minimum support
λ_j^k	Transition rate of component j out of state k
X_{FDEP}	Pattern of dependent alarms
I_{FDEP}	Metric of dependency

1 Introduction

The identification of Functional Dependencies (FDEPs) in Complex Technical Infrastructures (CTIs) has gained interest in the last years (Billinton and Allan 1992; Zio 2016; Serio et al. 2018; Rebello et al. 2018; Hickford et al. 2018; Antonello et al. 2019; Cantelmi et al. 2021). Given the CTIs complexity and evolutionary behaviour, the identification of FDEPs by classical methods of system decomposition and logic analysis is quite unattainable (Zio 2016; Rebello et al. 2018).

In small- and medium-scale systems, functionally dependent components or dependent abnormal behaviours are typically identified by analysing the system structure and the functional logic, considering design information and theoretical operation scenarios (Zio 2007).

For CTIs, some works have recently emphasized the importance of the identification of functionally dependent components or abnormal behaviours, which are typically unknown. General guidelines and conceptual definitions have been provided in Zio (2016). In this context, data-driven methods for the identification of FEDPs in CTIs using alarm data have been developed (Serio et al. 2018; Antonello et al. 2019; Antonello et al. 2021a). They are based on the application of the Association Rule Mining (ARM) (Agrawal and Imieliński 1993; Srikant and Agrawal 1996; Witten and Frank 2016) algorithm for scanning the alarm databases and identifying associations among patterns of alarms in the form of "if (antecedent) then (consequent)" rules; from these, the FDEPs are derived. Specifically, Apriori-based algorithms employ a level-wise iterative search mechanism, which scans the database to identify "frequent" patterns, and drives the search for other "frequent" patterns which contain the alarms of those patterns previously identified (Srikant and Agrawal 1996). A pattern is considered only if its frequency of occurrence is larger than a predefined threshold, called minimum support. Once a group of functionally dependent components has been identified, the causal chains of malfunctions can be obtained resorting to the knowledge of operators and experts of the CTI or by applying algorithms ad hoc developed to this aim. For example, a modified version of the quicksort algorithm has been developed in Antonello et al. (2020a) for the identification of the causal sequence of malfunctions from the probabilistic analysis of the temporal sequences of the alarms.

A main challenge in the application of the Apriori-based algorithms to alarm databases is the difficulty of identifying rare FDEPs, which are typically unknown and can be actually the most relevant for CTI vulnerability (Wang et al. 2000; Kim and Yun 2016; Zio 2016; Antonello et al. 2021a). Their identification typically requires the use of a small value for the minimum support threshold, which renders the search computationally unaffordable (Lin and Tseng 2006; Wulandari et al. 2019) and leads to the generation of a very large set of rules, which are not strongly supported and hard to analyse for discovering vulnerabilities in the CTI (Marin et al. 2008; Zhang et al. 2013). As a consequence, a relatively large value of minimum support threshold is employed at the risk of (i) not identifying rare patterns of alarms and (ii) extracting somewhat trivial association rules, which are already known to the CTI operators (Antonello et al. 2021a).

Multi-Objective Evolutionary Algorithms (MOEAs) have been recently proposed to directly identify association rules, also rare ones, eliminating the intermediate step of frequent-pattern mining and the related setting a minimum support threshold (Yan et al. 2009; Mukhopadhyay et al. 2014; Badhon et al. 2019). MOEAs are meta-heuristic approaches inspired by the laws of biological evolution, based on operations such as selection, recombination and mutation. The application of MOEAs for rule mining requires to evolve a population of candidate association rules according to properly defined rule metrics (Mukhopadhyay et al. 2014; Badhon et al. 2019). A limitation of the use of MOEAs for rule mining is the tendency of converging toward one or a limited set of optimal solutions, even though ARM applications usually require the identification of all the relevant rules (Martín et al. 2016).

Specific to the context of FDEPs identification in CTIs, an analyst is interested in identifying all the FDEPs influencing the CTI vulnerability (Antonello et al. 2021a). Thus, multiple solutions should be identified during the search and maintained in the population for effectively exploring the solution space and preventing premature convergence to local optima. In Antonello et al. (2020c), a MOEA has been developed for the identification of FDEPs in CTIs, employing the novelty search proposed in Lehman and Stanley (2011) to drive the exploration of the solutions space without being trapped in local optima. The novelty search drives the selective pressure to favour diversification in the population by dynamically rewarding the chromosomes based on their novelty with respect to other chromosomes, instead of rewarding them considering static fitness functions (Gomes et al 2017). While the approach allows the identification of

rare FDEPs in CTIs Antonello et al. (2021d), the following issues still need to be resolved:

(a) The MOEA tends to identify several rules including "spurious" alarms that have occurred by chance with other alarms, even if they do not belong to real FDEPs. Notice that the identification of patterns with spurious alarms can lead to possible errors when modelling FDEPs and cascading failures (Antonello et al. 2021c).
(b) The MOEA identifies several association rules derived by the same FDEP but differing for the combination of alarms in the antecedent and consequent parts. Considering a FDEP involving R alarms, the number of association rules which can be generated is $3^R - 2^{R+1} + 1$ (Del Jesus et al. 2011). Thus, when large alarm databases are considered, thousands (or tens of thousand) of associations rules are typically generated and, then, have to be post-processed to identify the relevant FDEPs, leading to a very large computational burden.

The present work extends the MOEA proposed in Antonello et al. (2020b) to address the above-mentioned issues. To this aim, the recently proposed metric of dependency (Antonello et al. 2021c), which has been shown to discriminate rules with spurious alarms from rules describing actual FDEPs (Antonello et al. 2021c), is used as fitness function within the MOEA search.

The main contributions of the proposed method are

- It allows discovering patterns of dependent alarms, without the preliminary step of identifying association rules;
- It allows discovering rare FDEPs and is robust with respect to spurious alarms occurring by chance at the same time of other real alarms;
- It incorporates for the first time in the MOEA the metric of dependency proposed in Antonello et al. (2020c); and
- It significantly reduces the computational burden required for the identification of rare FDEPs with respect to the Apriori-based algorithms.

The effectiveness of the proposed method is shown by means of its application to (i) an artificial case study, which mimics the complexity of a real CTI, and (ii) a real large-scale database of alarms generated by different supervision systems of the CTI of CERN, where a 27-km-perimeter ring particle accelerator composed by thousands of components is located.

The remainder of the paper is organized as follows: Sect. 2 describes the problem setting and the considered alarm database representation. In Sect. 3, the proposed MOEA is described. Section 4 introduces the case studies and discusses the obtained results. Finally, Sect. 5 draws some conclusions.

2 Problem setting

2.1 Alarm data representation

We consider a CTI formed by a large number of components, $N_c \gg 1$, and a database containing a large number of alarm messages,$N^{al} >> 1$, generated by the CTI during a long period of time $[t_0, t_f]$. The types of alarms associated to the generic j-th component, c_j, are M_j^{al}, and the total number of types of alarm messages is $M^{al} = \sum_{j=1}^{N_c} M_j^{al}$. The label $a_j^{k_j}$ refers to the k_j-th type of alarm message associated to the j-th component and $A = \left\{ a_1^1, \ldots, a_1^{M_1^{al}}, \ldots, a_j^1, \ldots, a_j^{M_j^{al}}, \ldots, a_{N_c}^1, \ldots a_{N_c}^{M_{N_c}^{al}} \right\}$ is the set of all the types of alarm messages.

Alarm messages are generated when the monitored signals exceed pre-set thresholds and are stored in the alarm message database (Fig. 1 a). The overall time interval $[t_0, t_f]$ is subdivided into Z consecutive small time intervals of the same length $\Delta t = \frac{t_f - t_0}{Z}$ (Fig. 1 (b)). A Boolean variable, $s_j^{k_j}(z)$, is associated to the occurrence of an alarm of type $a_j^{k_j}$ generated by component c_j in the z-th time interval:

$$s_j^{k_j}(z) = \begin{cases} 1 \text{ if alarm} a_j^{k_j} \text{occurs at least once in} \\ \quad [t_0 + (z-1)\cdot\Delta t, t_0 + z\Delta t] \\ 0 \text{ otherwise} \end{cases}, \tag{1}$$

and the state of the CTI during the generic z-th time interval is represented by the Boolean vector:

$$\vec{\mathcal{T}}(z) = \left[\begin{array}{c} s_1^1(z), \ldots, s_1^{M_1^{al}}(z), \ldots, s_j^1(z), \ldots \\ \ldots, s_{N_c}^1(z), \ldots s_{N_c}^{M_{N_c}^{al}}(z) \end{array} \right] \in [0,1]^{M^{al}}. \tag{2}$$

Finally, the database of alarms $(t_i, m_i), i = 1, \ldots, N^{al}$, is transformed into the Boolean matrix (Fig. 1 (b)):

$$T = \left[\begin{array}{c} \vec{\mathcal{T}}(1) \\ \ldots \\ \vec{\mathcal{T}}(Z) \end{array} \right] \in [0,1]^{Z \times M^{al}}, \tag{3}$$

whose generic z-th row represents the state of the CTI during the z-th time interval. Therefore, T provides a dynamic representation of the CTI state evolution in the time interval $[t_0, t_f]$.

2.2 Functional dependency (FDEP)

Two components are considered functionally dependent if the operation of one is influenced by the operation of the

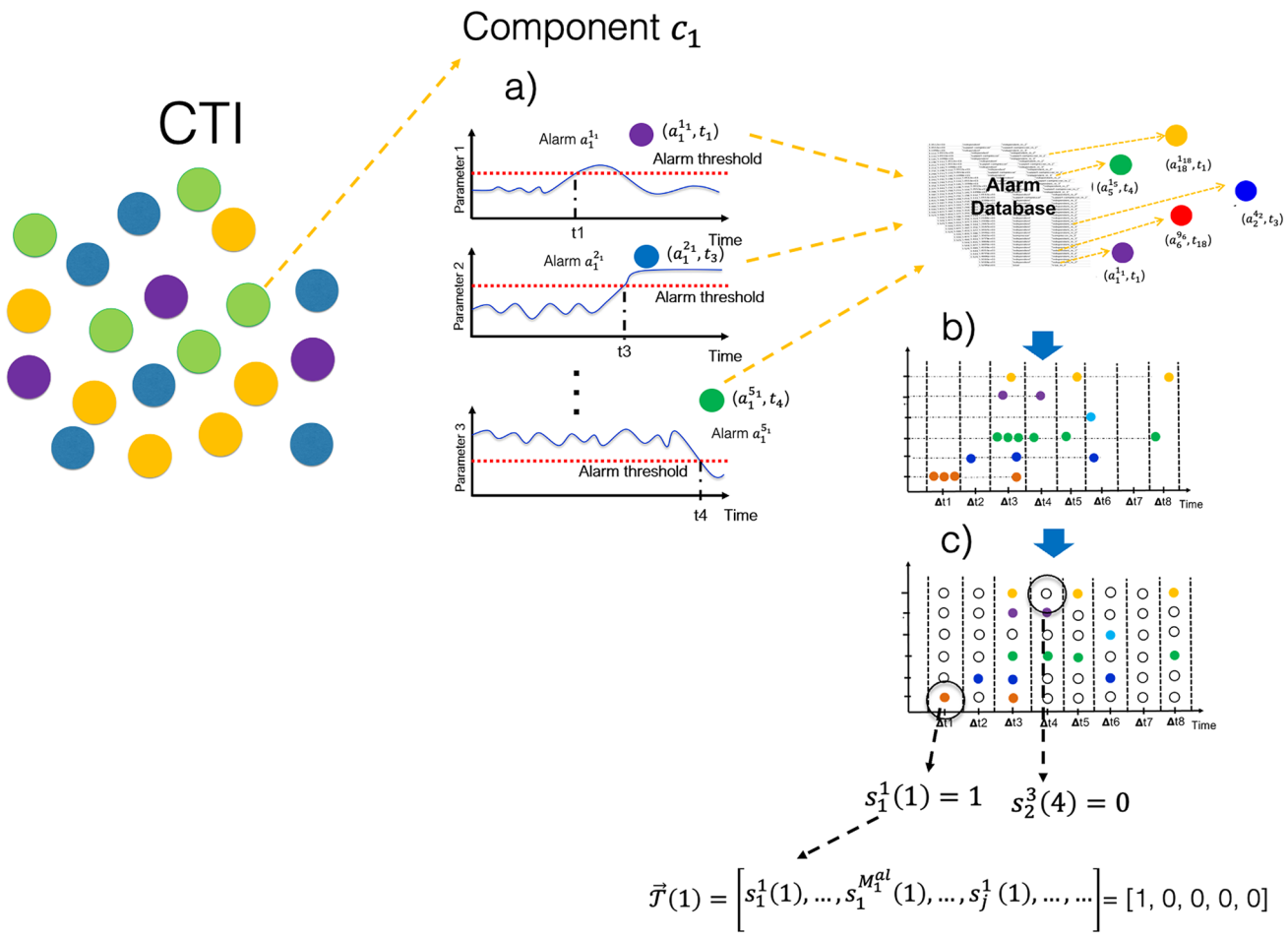

Fig. 1 **a** alarm messages generation and collection, **b** subdivision of the overall time interval $[t_0, t_f]$ into consecutive small time intervals, **c** representation of the alarm database using a Boolean matrix

other (Etesami and Kiyavash 2017). In particular, considering alarm messages, which are typically triggered when components have abnormal behaviours or malfunctions, we assume that there is a FDEP among two components of a CTI, c_1 and c_2, if a malfunction of component c_1, revealed by an alarm,$a_1^{k_1}$, causes a malfunction of components c_2, revealed by another alarm,$a_2^{k_2}$, or vice versa. This definition of FDEP assumes that the CTI monitoring system can detect all possible component malfunctions by measuring the proper physical quantities. In practice, some physical quantities related to rare or unknown component malfunctions are not monitored. As a consequence, FDEPs containing malfunctions not properly monitored cannot be identified by the analysis of the alarm messages.

2.3 A metric for the identification of FDEPs from alarm databases

Considering a generic pattern of $R \leq M^{al}$ alarms, $X = \left\{x_1, \ldots, x_j, \ldots, x_{M^{al}}\right\} \subseteq A$, the degree of dependency among the alarms $x_j \in X$, where $x_j = a_j^{k_j}$, for the sake of notation simplicity, can be assessed using the metric (Antonello et al. 2021c):

$$I_{\mathrm{FDEP}}(X) = \begin{cases} \dfrac{P(X)}{\prod_{x_j \in X} P(x_j)} & \text{if } P(X)\alpha \cdot \max_{x_j \in X}\left(P\left(x_j\right)\right) \\ 0 & \text{otherwise} \end{cases}, \tag{4}$$

where α is a parameter defined in the interval $[0, 1]$. The metric is based on the definition of functional dependency according to which the probability of occurrence of a pattern of R functionally dependent alarms, $X_{\mathrm{FDEP}} = \left\{a_{j^1}^{k^1}, \ldots, a_{j^R}^{k^R}\right\}$, is

$$P\left(X_{\mathrm{FDEP}}\right) > P(a_{j^1}^{k^1}) \bullet \cdots \bullet P(a_{j^R}^{k^R}), \tag{5}$$

and, therefore, the ratio

$$I_{X_{\text{FDEP}}} = \frac{P(X_{\text{FDEP}})}{\prod_{r=1}^{R} P(a_{j^r}^{k^r})}, \tag{6}$$

is larger than 1. The condition $P(X) > \alpha \bullet \max_{a_{j^r}^{k^r} \in X}\left(P\left(a_{j^r}^{k^r}\right)\right)$, where $\max_{a_j^{k_j} \in X}\left(P\left(a_j^{k_j}\right)\right)$ is the largest probability of occurrence among the probabilities of occurrence of the alarms of X, is motivated by the need of eliminating spurious alarms from the patterns. It derives from the following considerations:

(a) given a pattern of functionally dependent alarms, $X_{\text{FDEP}} = \left\{a_{j^1}^{k^1}, \ldots, a_{j^R}^{k^R}\right\} \subseteq A$, the probability of occurrence of any alarm $P(a_{j^r}^{k^r})$, $r = 1, \ldots, R$, can be decomposed into the sum of two contributions:

$$\mathrm{P}(a_{j^r}^{k^r}) = P_{X_{\text{FDEP}}}(a_{j^r}^{k^r}) + P_{Ind}(a_{j^r}^{k^r}), \tag{7}$$

where $P_{X_{\text{FDEP}}}(a_{j^r}^{k^r})$ and $P_{Ind}(a_{j^r}^{k^r})$ are the probabilities that $a_{j^r}^{k^r}$ occurs due to the FDEP and due to an event independent of the FDEP, respectively (Mosleh 1991; Zio 2009; O'Connor and Mosleh 2016);

(b) the probability of occurrence of a generic alarm $a_{j^r}^{k^r}$ due to a rare FDEP, $P_{X_{\text{FDEP}}}(a_{j^r}^{k^r})$, is expected to be close to the probability of occurrence of the whole pattern involved in the FDEP, $P(X_{\text{FDEP}})$: $P_{X_{\text{FDEP}}}(a_{j^r}^{k^r}) \cong P(X_{\text{FDEP}})$;

(c) the probability of occurrence of the pattern, $P(X_{\text{FDEP}})$, is expected to be larger than the probability of co-occurrence (by chance) of any alarm $a_{j^r}^{k^r}$ of the pattern, $r = 1, \ldots, R$, and a spurious (independent) alarm $a_{j^s}^{k^s}$, $P\left(X_{\text{FDEP}}\right) > P(a_{j^r}^{k^r}) \bullet P(a_{j^s}^{k^s})$.

Therefore, the necessary condition for a generic pattern of alarms, $X = \left\{a_{j^1}^{k^1}, \ldots, a_{j^R}^{k^R}\right\} \subseteq A$, to be functionally dependent is that the probability of occurrence of the pattern, $P(X)$, is greater than a fraction $\alpha \in [0, 1]$ of the probability of occurrence, $P(a_{j^r}^{k^r})$, of each alarm, $a_{j^r}^{k^r}$, of the pattern (Antonello et al. 2021c):

$$P(X) > \alpha \bullet P(a_{j^r}^{k^r}), \forall a_{j^r}^{k^r} \in X. \tag{8}$$

The setting of parameter α should consider the trade-off between the desiderata of identifying rare FDEPs, which are among the most interesting for CTI vulnerability analysis (Wang et al. 2000; Lee et al. 2005; Antonello et al. 2021c) and excluding spurious alarms. Considering Eq. 10, the use of a large α value would drive the search to discover frequent FDEPs, with the associated risk of not identifying rare FDEPs. On the opposite, some spurious patterns can be identified as actual FDEPs using small values of α. In this work, the parameter α is set equal to the value of 0.03, which has allowed the identification of rare FDEPs in two different CTIs (Antonello et al. 2020c). Also, the analysis reported in the same work shows that no spurious alarms are identified using values of α in the range [0.01; 0.08].

Notice that the probability of occurrence of a generic pattern $X = \left\{a_{j^1}^{k^1}, \ldots, a_{j^R}^{k^R}\right\}$ can be estimated from the alarm database using its *support*:

$$S(X) = n(X), \tag{9}$$

where $n(X)$ is the counter of the number of vectors $\vec{\mathcal{T}}(z)$ of the database $T = \begin{bmatrix} \vec{\mathcal{T}}(1) \\ \ldots \\ \vec{\mathcal{T}}(Z) \end{bmatrix}$ characterized by the occurrences of at least all the alarms of the pattern X (i.e. $\forall x_j \subset X, s_j(z) = 1$). Therefore, Eq. 4 becomes

$$I_{\text{FDEP}}(X) \begin{cases} \frac{S(X)}{\prod_{x_j \in X} P(x_j)} \text{if } S(X)\alpha \cdot \max_{x_j \in X}\left(S\left(x_j\right)\right) \\ 0 \qquad\qquad \text{otherwise} \end{cases} \tag{10}$$

2.4 Work objective

The objective of this work is the development of a method for the identification of FDEPs which satisfy the definition given in Sect. 2.3 using the alarm data introduced in Sect. 2.1.

3 Method

The problem described in Sect. 2 is addressed by developing a MOEA. Section 3.1 describes the encoding–decoding procedure adopted for representing FDEPs by means of chromosomes, Sect. 3.2 introduces the novelty search-based MOEA and Sect. 3.3 illustrates the search objectives, initial population and genetic operators (Sect. 3.3).

3.1 Chromosomes

A pattern of alarms is represented by a Boolean chromosome encoded in a vector of M^{al} bits, where each bit is associated to a specific alarm and is equal to 1 (0) when the corresponding alarm is present (is not present) in the pattern Del Jesus et al. (2011), Mukhopadhyay et al. (2014), Badhon et al. (2019). Figure 2 gives an example of chromosome decoding.

3.2 Novelty Search

The key idea of novelty search is to reward the divergence of a chromosome from those already in the population, instead of only considering the performance as evaluated by the chromosomes fitness functions (Lehman and Stanley 2011).

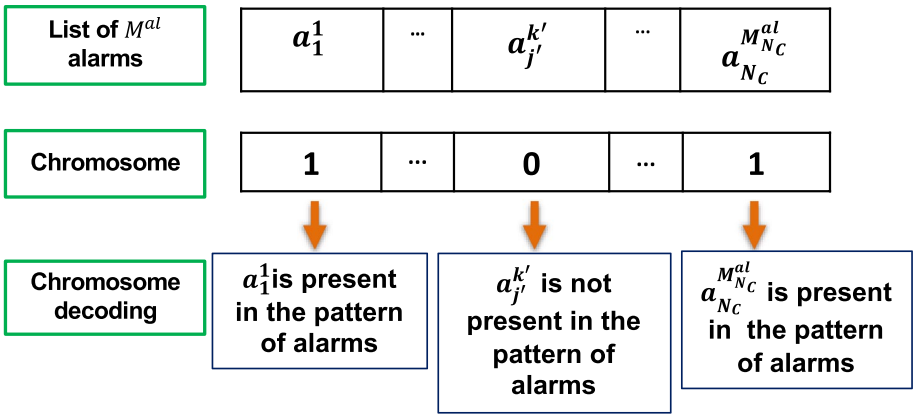

Fig. 2 Example of chromosome encoding/decoding

In practice, the uniqueness of a chromosome with respect to the rest of the population is evaluated introducing a metric of novelty, which measures the *sparseness* of the search space in correspondence of the chromosome as its average distance to the other chromosomes of the population. For a generic chromosome, ind_i, the metric of novelty is

$$\text{Novelty}\left(\text{ind}_i\right) = \frac{1}{N^{\text{pop}}} \sum_{j \neq i} \text{dist}(\text{ind}_i, \text{ind}_j), \tag{11}$$

where dist is a domain-dependent measure of the distance among chromosomes. The Jaccard distance $dist_J$, which has been shown to be effective in evaluating *sparseness* in pattern mining and Association Rule Mining applications (Tummala et al. 2018), is used here to evaluate the distance between a pair of chromosomes ind_i and ind_j

$$\text{dist}_J\left(\text{ind}_i, \text{ind}_j\right) = \text{dist}_J(X_i, X_j) = \frac{\left|X_i \cup X_j\right| - \left|X_i \cap X_j\right|}{\left|X_i \cup X_j\right|}, \tag{12}$$

where $|\,|$ refers to the cardinality of the pattern of alarms, i.e. the number of alarms contained in the pattern. Notice that, if the patterns X_i and X_j are identical, i.e. involve the same pattern of alarms, $\text{dist}_J\left(\text{ind}_i, \text{ind}_j\right)$ is equal to zero, otherwise, if the two patterns do not share any common alarm, $\text{dist}_J\left(\text{ind}_i, \text{ind}_j\right)$ is equal to 1. This metric favours the identification of novel patterns and allows comprehensively exploring the solution space by putting a constant evolutionary pressure on the search, but, at the same time, preserves unique and novel chromosomes (Lehman and Stanley 2011; Gomes et al. 2017).

3.3 MOEA algorithm

We use a Genetic Algorithm (GA) due to its straightforward principles, its simplicity of implementation and the fact that it has been already successfully applied to ARM and to other pattern mining problems (Anand et al. 2009; Badhon et al. 2019). As in Gomes et al. (2017), Antonello et al. (2020c), we combine novelty search and traditional fitness functions in the NSGA-II MOEA (Deb 2000), which is considered the most effective optimization algorithm for multi-objective rule mining. In this work, the objectives of the MOEA search are the maximization of (1) the novelty measure defined by Eq. (11), (2) the measure of dependency I_{FDEP} and (3) the *support* metric. Given the inclination of I_{FDEP} to favour rare patterns of alarms, as shown in Antonello et al. (2021c), the use of the *support* as third fitness function is needed for identifying frequent FDEPs as well. The combined use of these three objectives allows identifying patterns of dependent alarms while deeply exploring the solution space and avoiding premature convergence in local optima.

When one-bit genes are employed, the initialization of the population using sub-optimal chromosomes characterized by a limited number of one-bit genes equal to 1 can facilitate the GA convergence, as shown in the context of features selection problems (Baraldi et al. 2016) and association rules identification (Del Jesus et al. 2011; Antonello et al. 2020b). In this work, the initial population of chromosomes is created considering all the possible patterns, X', made of 2 alarms ($|X'| = 2$) which verify Eq. (8), and therefore, whose metric of dependency, $I_{\text{FDEP}}\left(X'\right)$, is larger than 0. Then, according to Antonello et al. (2020), we select the best N^{pop} chromosomes following the NSGA-II algorithm to set the initial population. This is consistent with the objective of the search, which is the identification of FDEPs, given that by selecting only the patterns which satisfy Eq. (8), we a priori discard the patterns made of 2 alarms which do not belong to FDEPs.

As suggested by Del Jesus et al. (2011), Mukhopadhyay et al. (2014) in case of standard binary chromosomes, standard genetic operators and an evolution algorithm based on a traditional GA with two-point crossover and flip mutation are used. Furthermore, we avoid the presence of identical chromosomes in the population to favour population diversity (Antonello et al. 2020b).

Table 1 Transition rates (hours^{-1})

Component c_j	Transition rates			
$j = 1, \ldots, 100$	$\lambda_j^{1\to2} = 0.5$	$\lambda_j^{2\to3} = 0.05$	$\lambda_j^{3\to4} = 0.07$	$\lambda_j^{4\to5} = 0.03$
	$\lambda_j^{2\to1} = 0.5$	$\lambda_j^{3\to2} = 0.01$	$\lambda_j^{4\to3} = 0.03$	$\lambda_j^{5\to1} = 0.2$
$j = 101, \ldots, 200$	$\lambda_j^{1\to2} = 0.2$	$\lambda_j^{2\to3} = 0.007$	$\lambda_j^{3\to4} = 0.01$	$\lambda_j^{4\to5} = 0.05$
	$\lambda_j^{2\to1} = 0.2$	$\lambda_j^{3\to2} = 0.01$	$\lambda_j^{4\to3} = 0.04$	$\lambda_j^{5\to1} = 0.2$
$j = 201, \ldots, 300$	$\lambda_j^{1\to2} = 0.1$	$\lambda_j^{2\to3} = 0.004$	$\lambda_j^{3\to4} = 0.04$	$\lambda_j^{4\to5} = 0.001$
	$\lambda_j^{2\to1} = 0.1$	$\lambda_j^{3\to2} = 0.01$	$\lambda_j^{4\to3} = 0.04$	$\lambda_j^{5\to1} = 0.2$

alarm $a_j^1, j = 1, \ldots, N_c$, is triggered when component c_j performs the transition from D=2 to D=3 and the alarm a_j^2 when it performs the transition from D=3 to D=4, whereas all the other transitions do not generate alarms.

We further assume the existence of seven different FDEPs (Table 2) among CTI components. They are originated by the transition from state 2 ('partially degraded') to state 3 ('degraded') of a component, which can cause the transition of an ordered sequence of components from state 4 to state 5. The probability of propagation of the functional dependencies from a component to the successive one of the chain

Table 2 Simulated FDEPs

No	Involved components	Trigged alarms	Probability of Propagation	Number of occurrences of the overall sequence in the simulated database	Number of occurrences of the overall sequence in the simulated database in a single time interval
1	$c_{1,1}, c_{2,1}, c_{3,1}\, c_{4,1}, c_{5,1}, c_{6,1}, c_{7,1}$ $c_{8,1}, c_{9,1}, c_{10,1}$	$a^1_{1,1} \to a^2_{2,1} \to a^2_{3,1} \to a^2_{4,1} \to a^2_{5,1} \to a^2_{6,1} \to a^2_{7,1} \to a^2_{8,1} \to a^2_{9,1} \to a^2_{10,1}$	0.9	119	113
2	$c_{1,2}, c_{2,2}, c_{3,2}\, c_{4,2}, c_{5,2}, c_{6,2}, c_{7,2}$ $c_{8,2}, c_{9,2}, c_{10,2}$	$a^1_{1,2} \to a^2_{2,2} \to a^2_{3,2} \to a^2_{4,2} \to a^2_{5,2} \to a^2_{6,2} \to a^2_{7,2} \to a^2_{8,2} \to a^2_{9,2} \to a^2_{10,2}$	0.9	41	37
3	$c_{1,3}, c_{2,3}, c_{3,3}\, c_{4,3}, c_{5,3}, c_{6,3}, c_{7,3}$ $c_{8,3}, c_{9,3}, c_{10,3}$	$a^1_{1,3} \to a^2_{2,3} \to a^2_{3,3} \to a^2_{4,3} \to a^2_{5,3} \to a^2_{6,3} \to a^2_{7,3} \to a^2_{8,3} \to a^2_{9,3} \to a^2_{10,3}$	0.9	15	13
4	$c_{11,1}, c_{12,1}, c_{13,1}, c_{11,2}, c_{12,2}$ $, c_{13,2}, c_{11,3}, c_{12,3}, c_{13,3}$	$a^1_{11,1} \to a^2_{12,1} \to a^2_{13,1} \to a^2_{11,2} \to a^2_{12,2} \to a^2_{13,2} \to a^2_{11,3} \to a^2_{12,3} \to a^2_{13,3}$	0.95	34	34
5	$c_{21,1}, c_{22,1}, c_{21,3}, c_{22,3}$	$a^1_{21,1} \to a^2_{22,1} \to a^2_{21,3} \to a^2_{22,3}$	0.85	34	33
6	$c_{21,2}, c_{22,2}, c_{31,3}, c_{33,3}$	$a^2_{21,2} \to a^2_{22,2} \to a^2_{31,3} \to a^2_{33,3}$	0.95	13	10
7	$c_{31,2}, c_{32,2}, c_{41,3}, c_{42,3}$	$a^2_{31,2} \to a^2_{32,2} \to a^2_{41,3} \to a^2_{42,3}$	0.95	10	8

4 Case studies

The proposed method is applied to a synthetic database of alarms generated by simulating the behaviour of a CTI and to a database of real alarms generated by the technical infrastructure of CERN during one year of operation.

4.1 Synthetic alarm database

We consider the alarm database of (Antonello et al. 2021b). It contains the alarm messages generated by the simulation of a CTI formed by N_c=300 components, each of which can be in five mutually exclusive and exhaustive states $D \in \{1, 2, 3, 4, 5\}$ corresponding to healthy ($D = 1$), partially degraded ($D = 2$), degraded ($D = 4$), very degraded ($D = 4$) and failed ($D = 5$) conditions. A generic component $c_j, j = 1, \ldots, N_c$, performs transitions among the states at exponentially distributed random times. Table 1 reports the constant transition rates among the different states. The is reported in Table 2. Notice that functional dependencies No 6 and No 7 are the rarest, since their initiation depends on the failure of components characterized by low probabilities of failure.

The CTI behaviour has been simulated for a period of time $[t_0, t_f] = [0,\ 365\ \text{days}]$ during which 172,225 alarm messages reporting $M^{al} = 600$ different types of malfunctions have been generated by the 300 CTI components. The fifth column of Table 2 reports the number of occurrences of the overall sequences of dependent alarms, which range from 8 to 119. The entire time domain of the analysis is discretized into $Z = 8760$ time intervals of length $\Delta t = 60$ min. As reported in the last column, this setting of Δt guarantees that the whole sequence of alarms occurs into a single time step in a significant fraction of times (Antonello et al. 2021a). The last column of Table 2 reports the number of occurrences in the database of the chain of alarms in a single time interval of 60 min. As expected, the whole chain of alarms occurs in a single

Table 3 Functional dependencies and corresponding identified patterns of alarms

Simulated functional dependence		Extracted patterns			
	Involved alarms	No.	Pattern	$\log(I_{\mathrm{FDEP}})$	Support
FEDP 1	$a_1^1 \rightarrow a_2^2 \rightarrow a_3^2 \rightarrow a_4^2 \rightarrow a_{5,1}^2 \rightarrow a_6^2 \rightarrow a_7^2 \rightarrow a_8^2 \rightarrow a_9^2 \rightarrow a_{10}^2$	1	$\left\{ \begin{matrix} a_1^1, a_2^2, a_3^2, a_4^2, a_5^2, a_6^2, a_7^2, a_8^2, \\ a_9^2, a_{10}^2 \end{matrix} \right\}$	33.15	113
		2	$\{a_1^1, a_2^2, a_3^2, a_4^2, a_5^2, a_6^2, a_7^2, a_9^2, \}$	26.37	140
		3	$\left\{ \begin{matrix} a_1^1, a_2^2, a_6^2, a_7^2, a_8^2, \\ a_9^2, a_{10}^2 \end{matrix} \right\}$	25.83	145
		4	$\left\{ \begin{matrix} , a_4^2, a_5^2, a_6^2, a_7^2, a_8^2, \\ a_9^2, a_{10}^2 \end{matrix} \right\}$	23.55	147
		5	$\left\{ \begin{matrix} a_1^1, a_2^2, a_3^2, a_4^2, , a_7^2, a_8^2, \\ a_{10}^2 \end{matrix} \right\}$	21.15	155
			…		
FEDP 2	$a_{120}^1 \rightarrow a_{121}^2 \rightarrow a_{122}^2 \rightarrow a_{123}^2 \rightarrow a_{124}^2 \rightarrow a_{125}^2 \rightarrow a_{126}^2 \rightarrow a_{127}^2 \rightarrow a_{128}^2 \rightarrow a_{129}^2$	130	$\{a_{120}^1, a_{121}^2, a_{122}^2, a_{123}^2, a_{124}^2, a_{125}^2, a_{128}^2, a_{129}^2, a_{126}^2, a_{127}^2\}$	44.61	37
		131	$\{a_{126}^2, a_{127}^2, a_{128}^2, a_{129}^2\}$	18.65	*55*
			…		
FEDP 3	$a_{230}^1 \rightarrow a_{231}^2 \rightarrow a_{232}^2 \rightarrow a_{233}^2 \rightarrow a_{234}^2 \rightarrow a_{235}^2 \rightarrow a_{236}^2 \rightarrow a_{237}^2 \rightarrow a_{238}^2 \rightarrow a_{239}^2$	210	$\left\{ \begin{matrix} a_{239}^2, a_{230}^1, a_{231}^2, a_{232}^2, a_{234}^2, \\ a_{235}^2, a_{236}^2, a_{237}^2, a_{238}^2, a_{233}^2 \end{matrix} \right\}$	53.13	13
		211	$\left\{ \begin{matrix} a_{230}^1, a_{231}^2, a_{232}^2, a_{234}^2, \\ a_{235}^2 \end{matrix} \right\}$	37.02	20
			…		
FEDP 4	$a_{11}^1 \rightarrow a_{12}^2 \rightarrow a_{13}^2 \rightarrow a_{111}^2 \rightarrow a_{112}^2 \rightarrow a_{113}^2 \rightarrow a_{211}^2 \rightarrow a_{212}^2 \rightarrow a_{213}^2$	306	$\{a_{113}^2, a_{112}^2, a_{12}^2, a_{212}^2, a_{11}^2, a_{213}^2, a_{111}^1, a_{211}^2, a_{13}^2\}$	38.88	34
		307	$\{a_{131}^2, a_{112}^2, a_{122}^2\}$	12.61	*50*
			…		
FEDP 5	$a_{201}^1 \rightarrow a_{202}^2 \rightarrow a_{203}^2 \rightarrow a_{204}^2$	401	$\left\{ a_{201}^1, a_{202}^2, a_{204}^2, a_{203}^2 \right\}$	14.14	33
			$\{a_{201}^1, a_{202}^2\}$	6.74	*35*
			…		
FEDP6	$a_{21}^2 \rightarrow a_{22}^2 \rightarrow a_{31}^2 \rightarrow a_{33}^2$	520	$\{a_{21}^2, a_{22}^2, a_{33}^2, a_{32}^2\}$	18.43	11
		421	$\{a_{33}^2, a_{21}^2\}$	7.18	15
			…		
FEDP 7	$a_{131}^2 \rightarrow a_{132}^2 \rightarrow a_{241}^2 \rightarrow a_{242}^2$	450	$\{a_{132}^2, a_{241}^2, a_{242}^2, a_{131}^2\}$	17.72	8
		551	$\{a_{131}^2, a_{241}^2\}$	7.97	12
			…		

time interval in most of the cases. For example, functional dependency no 7, whose expected propagation time is 10 min, occurs 10 times and all alarms of the FDEP occur in the same time interval in 8 cases.

The proposed MOEA algorithm is applied using an initial population of $N^{\mathrm{pop}} = 500$ binary chromosomes of M^{al} bits. The mutation probability is set to $1/M^{\mathrm{al}}$ and the crossover probability to 0.8 according to (Anand et al. 2009). The algorithm has been run for 2000 generations, obtaining a final set of 500 patterns in a computational time of 657 s on an Intel core (TM) i7-4790 CPU@ 3.6 GHz, 16 GB RAM. The final population includes several patterns involving the alarms of each of the FDEPs of Table 2. Table 3 gives, for each of the FDEPs of Table 2, the pattern in the final population with the largest I_{FDEP} value.

Table 3 reports some examples of patterns of the final population containing groups of alarms which belong to the FDEPs of Table 2. Notice that many patterns only partially describe the FDEPs, i.e. they do not contain all the involved alarms. For example, pattern nos. 2, 3, 4 and 5 contain only 8 of the 10 alarms of FDEP 1. On the other side, it is interesting to observe that the pattern with the largest I_{FDEP} value always contains all the alarms of the corresponding FDEP. This highlights the capability of the metric of dependency to identify the pattern formed by all the alarms of the FDEP, which simplifies the post-processing of the results. Also, the analysis of all the patterns of the final population has

Table 4 Results of the proposed MOEA considering different combinations of the fitness functions

Search objectives	Computational time (on Intel core (TM) i7-4790 CPU@ 3.6 GHz, 16 GB RAM)	Number of identified FDEPs	Identifications of patterns with spurious alarms
I_{FDEP}, support and novelty	657 s	7	No
I_{FDEP} and novelty	623 s	5	No
Support and novelty	619 s	1	Yes
I_{FDEP} and support	612 s	4	No
I_{FDEP}	607 s	1	No

Table 5 Results comparison among the proposed MOEA, the MOEA for association rules identification proposed in Antonello et al. (2020) and the Apriori-based ARM algorithm presented in Antonello et al. (2021)

Approach	Computational time (on an Intel core (TM) i7-4790 CPU@ 3.6 GHz, 16 GB RAM)	Patterns identified	Functional dependencies identified	Presence of spurious rules
Proposed novelty-based MOEA	11 min	500	7	No
Apriori-based ARM algorithm	6271 min	2000	7	Yes
MOEA for association rules identification	14 min	500	7	Yes

shown that none of them contains spurious alarms. The post-processing procedure for the identification of the FDEPs from the patterns of the final population requires to (1) sort them with respect to I_{FDEP} and (2) eliminate the patterns containing subsets of alarms already contained in patterns with larger I_{FDEP}.

Table 4 reports the results of the proposed MOEA on the same database considering different combinations of fitness functions. As expected, when novelty search is not used, the final population converges to a population of patterns describing only few functional dependencies. Also, the use of the metric of dependency I_{FDEP} guarantees that the identified patterns do not contain spurious alarms.

The results obtained by the proposed novelty-based MOEA have been compared with the results of the Apriori-based ARM algorithm proposed in Antonello et al. (2021) and of the MOEA for ARM identification proposed in Antonello et al. (2020). The Apriori-based ARM algorithm performs an exhaustive search among all the possible combinations of alarms but requires to set small values of the minimum support and minimum confidence thresholds (here chosen equal to 5 and 0.6, respectively) to identify all the rare FDEPs; otherwise, with larger values of these thresholds, it would not find them (Antonello et al. 2021b). The MOEA for ARM identification evolves a population of 500 association rules encoded in binary chromosomes of $2 \times M^{al}$ bits and employs the novelty measure (Eq. 9) and the metrics of *Interestingness* and *Length* (Pachón Álvarez and Mata Vázquez 2012; Dhaenens and Jourdan 2016) as search objectives.

Table 5 reports the obtained results. The Apriori-based ARM algorithm requires a computational effort more than 500 times larger than the approach proposed in this work and produces 2000 different association rules, which must be post-processed to discriminate the rules containing spurious alarms and to identify the rare FDEPs of particular interest for vulnerability analysis. The MOEA for ARM identification finds 500 association rules, which contain all the 7 FDEPs in a computational time slightly larger than the proposed approach. A limitation of this approach is that 15% of the generated association rules contain spurious alarms, which requires the identified rules to be analysed one by one by plant experts in order to distinguish the actual FDEPs.

To conclude, the comparison has shown that the proposed MOEA is able to *i*) correctly identify the functional dependencies with a reduction of the computational effort with respect to the other two approaches considered; *ii*) discover rare FDEPs without requiring the setting of a very low value for minimum support; and *iii*) be robust against spurious alarms.

4.2 CERN complex technical infrastructure

The CTI of CERN is composed by several systems working together to support the operation of the LHC, which is the largest existing particle accelerator in the World (Nielsen and Serio 2016). It consists of a 27-km ring of superconducting magnets and infrastructures, extending over the Swiss and French borders and located about 100 m underground.

A database of alarms generated during the period $[t_0, t_f] = [01$ January 2016; 31 December 2016] by three

Table 6 Example of patterns of alarms generated by components of different systems

	Pattern	I_{FDEP}	Support
1	['SU_8_UPEA802_AL6.IST', 'SU_8_UPKA802_AL6.IST', 'EKD202_SLASH_8U_I1314']	9.56	53
2	['QSRB_8_PV100_SI1.IST', 'QSRB_8_PV200_SI1.IST', 'QSRB_8_PV279_SI1.IST', 'QSRB_8_PV249_SI1.IST', 'QSRB_8_CV003_FS1.IST', 'QSV_8_BEEPFS81.IST', 'EKD104_SLASH_8HM_I1314']	38.75	3
3	['QSRB_8_PV100_SI1.IST', 'QSRB_8_PV279_SI1.IST', 'QSRB_8_PV249_SI1.IST', 'QSRB_8_CV003_FS1.IST', 'EKD104_SLASH_8HM_I1314']	27.02	3
4	['QSRB_8_PV200_SI1.IST', 'QSRB_8_PV279_SI1.IST', 'QSRB_8_PV249_SI1.IST', 'QSRB_8_CV003_FS1.IST', 'EKD104_SLASH_8HM_I1314']	26.13	3
5	['QSCA_8_CSC1_SI3.IST', 'QSCA_8_CSC1_SI6.IST', 'EKD210_SLASH_8U_S3S16']	12.16	11

supervision systems of a representative part of the LHC infrastructure, the LHC *point 8*, is considered. During this period, N_{al} = 18,711,737 alarms reporting M^{al}=13,451 different types of malfunctions have occurred. The alarm database has been pre-processed by pruning those alarms involved in already-known FDEPs among components belonging to a same system and which are, therefore, less interesting from the point of view of the system vulnerability (Antonello et al. 2020b). The pruned database consists in N_{al} = 112,591 alarm messages reporting M^{al}=1024 different types of malfunctions.

Considering the expected time of propagation of a FDEP, which is influenced by the physical characteristics of the systems and processes involved, the time interval for the analysis is set equal to $\Delta t = 30$ min to ensure the identification of all FDEPs while minimizing computational resources and spurious alarms (Antonello et al. 2021). Therefore, the one-year period [01 January 2016; 31 December 2016] is divided into $Z = 17{,}500$ time intervals. Setting the mutation probability equal to $1/M^{al}$ and the crossover probability equal to 0.8 (Anand et al. 2009), an initial population of $N^{pop} = 500$ individuals, encoded into chromosomes of M^{al} bits, is evolved for 2000 generations obtaining a final set of 500 patterns of alarms in 641 s on an Intel core (TM) i7-4790 CPU@ 3.6 GHz, 16 GB RAM. The patterns describe FDEPs involving components of different systems, whose failures can cause a local malfunction to propagate across the CTI systems and sub-systems, and originate unexpected cascades of failures over vast geographic areas (Thacker et al. 2017; Antonello et al. 2021b). Table 6 reports a selection of five discovered patterns which have been considered as most representative example of novel and unknown chains of events by CERN experts. The first pattern describes the correlation among malfunctions involving a breaker of the electric system, powering the cryogenic system (EKD202_SLASH_8U) and malfunctions of two pumps of the cooling and ventilation system (SU_8_UPKA802_AL6, SU_8_UPEA802_AL6). The second pattern describes the associated occurrence of a malfunction in a breaker of the electrical system (EKD204_SLASH_8U) and three different malfunctions in the helium compressors of the cryogenic system (QSCB_8_CSY_C2). Pattern 3 and Pattern 4 describe the propagation of a malfunction triggered by problems in the cryogenic electric system distribution switchboard ('EKD104_SLASH_8HM_I1314', 'EKD107_SLASH_8HM_I1314'), which propagate and lead to malfunctions of the Cryogenic system helium refrigerator, dryer and compressors ('QSAB_8_QSA_TS3.IST', 'QSCB_8_CSY_C1_SI3.IST', 'QSRB_8_CV003_FS1.IST'). Pattern 5 describes the propagation of a malfunction of the electric system 'EKD210_SLASH_8U_S3S16' to the Cryogenic system helium compressors ('QSCA_8_CSC1_SI3.IST', 'QSCA_8_CSC1_SI6.IST'). Patterns 2, 3 and 4 can be considered as rare since the corresponding alarms occur in the same time interval only three times in a period of one year, during which a whole of 112591alarms has been generated. Notice that patterns with support smaller than 3 are not identified by the algorithm due to the need of avoiding spurious FDEPs, which satisfy Eq. 10 by chance Antonello et al. (2021c).

An independent expert analysis has confirmed that the involved components are, indeed, part of chains of malfunctions that occurred in 2016. According to the CTI experts, the identification of rare functional dependencies is useful for (i) updating the maintenance plan of the components involved in the functional dependencies, for example, by increasing the frequency of inspection for those components that cause the chains of events with the objective of reducing the probability of their initiation; (ii) upgrading the most critical components; (iii) introducing barriers to contrast the propagation of the chain of events; and (iv) facilitating root cause analysis.

To further analyse the advantages of the proposed evolutionary approach, a traditional Apriori-based algorithm (Antonello et al. 2019) is applied to the same database. In order to be able to identify rare rules, the value of *minimum support* threshold is now set equal to 3 and the *minimum confidence* is set equal to 0.6%. The search has produced 1049 association rules in a computational time of 43,959 s on an Intel core (TM) i7-4790 CPU@ 3.6 GHz, 16 GB RAM. Notice that the proposed MOEA allows reducing the

computational effort (641 s) with respect to traditional ARM (43,959 s) and does not require the setting of a minimum support.

5 Conclusions

Functional Dependencies (FDEP) in Complex Technical Infrastructures (CTIs) need to be identified to analyse potential vulnerabilities. This work proposes a MOEA based on novelty search and on a recently proposed metric of dependency for the identification of FDEPs from alarm data.

A main novelty with respect to the other state-of-the-art approaches is the direct identification of patterns of dependent alarms, without the preliminary step of identifying association rules. This avoids setting minimum support and minimum confidence thresholds and allows swiftly disclosing also rare FDEPs. The novelty metric drives the search to favour diversification in the results, allows to explore the solution space and avoids to be trapped in local optima. Moreover, the use of the metric of dependency allows to discriminate spurious alarms and, therefore, eliminates the results post-processing step and reduces the computational burden.

An application to a synthetic database of alarms has shown the ability of the proposed MOEA to effectively explore the solution space for identifying all actual (i.e. not spurious) simulated functional dependencies. Comparison with an Apriori-based ARM algorithm and a MOEA for association rules identification shows (i) the ability of the proposed approach to be robust to spurious alarms in the FDEPs, (ii) the low computational effort required by the proposed approach and (iii) the reduction in the number of redundant or not completely identified patterns of FDEPs found by the proposed approach, which would complicate the post-processing of the results.

The application of the proposed algorithm to a large-scale database collected at CERN CTI has allowed identifying patterns of alarms which describe unknown and rare FDEPs in the CTI, which have then been confirmed by CERN experts as indeed responsible of sequences of malfunctions occurred in the past.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Agrawal R, Imieliński T (1993) Mining association rules between sets of items in large databases. ACM SIGMOD Rec 22(2):207–216

Anand R, Vaid A, Singh PK (2009) Association rule mining using multi-objective evolutionary algorithms: strengths and challenges. In: World congress on nature and biologically inspired computing, NABIC 2009—proceedings 5393878, pp 385–390

Antonello F, Baraldi P, Shokry A, Zio E, Gentile U, Serio L (2020a) Data-driven extraction of association rules of dependent abnormal behaviour groups. In: Proceedings of the 29th European safety and reliability conference, ESREL 2019, pp 3308–3313

Antonello F, Baraldi PA, Gentile U, Serio L, Shokry A, Zio E (2020b) Multi-objective evolutionary algorithm for the identification of rare functional dependencies in complex technical infrastructures. In: Proceedings of the 30th European safety and reliability conference and the 15th probabilistic safety assessment and management conference, pp. 325–333

Antonello F, Baraldi PA,Gentile U, Serio L, Shokry A, Zio E (2020c) A method for inferring causal dependencies among abnormal behaviours of components in complex technical infrastructures. In: Proceedings of the 30th European safety and reliability conference and the 15th probabilistic safety assessment and management conference, pp 209–216.

Antonello F, Baraldi P, Shokry A, Zio E, Gentile U, Serio L (2021a) Association rules extraction for the identification of functional dependencies in complex technical infrastructures. Reliab Eng Syst Saf 209:107305

Antonello F, Baraldi PU, Serio L, Zio E (2021b) A novel association rule mining method for the identification of rare functional dependencies in complex technical infrastructures from alarm data. Expert Syst Appl 170:114560

Antonello F, Baraldi PU, Serio L, Zio E (2021c) A novel metric to evaluate the association rules for identification of functional dependencies in complex technical infrastructures. Environ Syst Decis 192:3134–3143

Antonello F, Baraldi P, Serio L, Zio E (2021d) A novelty search-based evolutionary algorithm for the identification of rare functional dependencies in complex technical infrastructures from alarm data. In: Proceedings of the 30th international European safety and reliability conference, ESREL-PSAM 2020

Badhon B, Kabir MMJ, Xu S, Kabir M (2019) A survey on association rule mining based on evolutionary algorithms. Int J Comput Appl 43(3):775–785

Baraldi P, Cannarile F, Di Maio F, Zio E (2016) Hierarchical k-nearest neighbours classification and binary differential evolution for fault diagnostics of automotive bearings operating under variable conditions. Eng Appl Artif Intell 56:1–13

Billinton R, Allan RN (1992) Network modelling and evaluation of complex systems. In: Reliability evaluation of engineering systems. Springer, Boston

Cantelmi R, Di Gravio G, Patriarca R (2021) Reviewing qualitative research approaches in the context of critical infrastructure resilience. Environ Syst Decis 24:1–36

Deb K, Agrawal S, Pratap A, Meyarivan T (2000) A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II. In: Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics), pp. 849–858

Dhaenens C, Jourdan L (2016) Metaheuristics for big data. Metaheuristics Big Data 5:1–188

Etesami J, Kiyavash N (2017) Measuring causal relationships in dynamical systems through recovery of functional dependencies. IEEE Trans Signal Inform Process over Netw 3(4):650–659

Gomes J, Mariano P, Christensen AL (2017) Novelty-driven cooperative coevolution. Evol Comput 25(2):275–307

Hickford AJ, Blainey SP, Ortega Hortelano A et al (2018) Resilience engineering: theory and practice in interdependent infrastructure systems. Environ Syst Decis 38:278–291

Del Jesus MJ, Gámez JA, González P, Puerta JM (2011) On the discovery of association rules by means of evolutionary algorithms. Wiley Interdiscip Rev 1(5):397–415

Lee YC, Hong TP, Lin WY (2005) Mining association rules with multiple minimum supports using maximum constraints. Int J Approx Reason 40(1–2):44–54

Lehman J, Stanley KO (2011) Novelty search and the problem with objectives. In: Riolo R, Vladislavleva E, Moore J (eds) Genetic programming theory and practice IX. Genetic and evolutionary computation. Springer, New York

Lin WY, Tseng MC (2006) Automated support specification for efficient mining of interesting association rules. J Inform Sci 32(3):238–250

Marin N, Molina C, Serrano JM, Vila MAA (2008) Complexity guided algorithm for association rule extraction on fuzzy datacubes. IEEE Trans Fuzzy Syst 16(3):693–714

Martín D, Alcalá-Fdez J, Rosete A, Herrera F (2016) NICGAR: a niching genetic algorithm to mine a diverse set of interesting quantitative association rules. Inform Sci 355:208–228

Mukhopadhyay A, Maulik U, Bandyopadhyay S, Coello CAC (2014) Survey of multiobjective evolutionary algorithms for data mining: part II. IEEE Trans Evol Comput 18(1):25–35

Pachón Álvarez V, Mata Vázquez J (2012) An evolutionary algorithm to discover quantitative association rules from huge databases without the need for an a priori discretization. Expert Syst Appl 39(1):585–593

Rebello S, Yu H, Ma L (2018) An integrated approach for system functional reliability assessment using dynamic Bayesian network and Hidden Markov model. Reliab Eng Syst Saf 180:124–135

Serio L, Antonello F, Baraldi P, Castellano A, Gentile U, Zio E (2018) Smart framework for the availability and reliability assessment and management of accelerators technical facilities. In: 9th International Particle Accelerator Conference, IPAC 2018

Thacker S, Pant R, Hall JW (2017) System-of-systems formulation and disruption analysis for multi-scale critical national infrastructures. Reliab Eng Syst Saf 167:30–41

Tummala K, Oswald C, Sivaselvan BA (2018) Frequent and rare itemset mining approach to transaction clustering. Commun Comput Inform Sci 804:8–18

Wang K, He Y, Han J (2000) Mining frequent itemsets using support constraints. In Proc. VLDB. p 43–52

Witten IH, Frank E (2016) Data mining: practical machine learning tool and techniques. Morgan Publishers, Burlington

Wulandari CP, Ou-Yang C, Wang H-C (2019) Applying mutual information for discretization to support the discovery of rare-unusual association rule in cerebrovascular examination database. Expert Syst Appl 118:52–64

Yan X, Zhang C, Zhang S (2009) Genetic algorithm-based strategy for identifying association rules without specifying actual minimum support. Expert Syst Appl 36(2):3066–3076

Zhang J, Wang Y, Feng J (2013) Attribute index and uniform design based multiobjective association rule mining with evolutionary algorithm. Sci World J 13:259347

Zio E (2007) An introduction to the basics of reliability and risk analysis. Series on quality, reliability and engineering statistics

Zio E (2009) Computational methods for reliability and risk analysis. World Scientific

Zio E (2016) Challenges in the vulnerability and risk analysis of critical infrastructures. Reliab Eng Syst Saf 152:137–150