

Construction of a Reliability Structure Function Based on Uncertain Data

Elena Zaitseva, and Vitaly Levashenko

Abstract— Structure function is one of the possible mathematical models of the real systems under study in reliability engineering. The structure function represents correlation the system performance level and components states. The system performance level is defined from the states of all its components. It means that all possible components states and performance levels must be indicated and reflected in the structure function. Initial data for the analysis of real system is uncertain. New methods to construct a structure function based on initial uncertain data is proposed. Fuzzy Decision Trees are used in this method to transform initial uncertain data about a real system into an exact-defined a system structure function. The proposed method includes three principal steps for the structure function construction: (a) collection of data in the repository; (b) representation of the system model in the form of an FDT; (c) construction of the structure function based on the FDT.

Index Terms— Multi-State System, Structure Function, Fuzzy Decision Tree, uncertainty

NOTATION

$\phi(\mathbf{x})$	structure function that defines system performance levels from failure ($\phi(\mathbf{x}) = 0$) to perfect functioning ($\phi(\mathbf{x}) = M - 1$);
M	number of the system performance levels
n	number of the system components
x_i	the i -th component state that changes from failure ($x_i = 0$) to perfect functioning ($x_i = m_i - 1$)
m_i	number of the i -th component states
$\mathbf{x} = (x_1, \dots, x_n)$	state vector
$p_{i,s}$	probability of state s of the i -th component
A_i	the i -th input attribute ($i = 1, \dots, n$)
$A_{i,j}$	value j of the i -th input attribute ($j = 0, \dots, m_i - 1$)
$\{A_{i,0}, \dots, A_{i,j}, \dots, A_{i,m_i-1}\}$	possible values of input attribute A_i
B	output attribute
B_s	value s of the output attribute ($s = 0, \dots, M - 1$)
$\{B_0, \dots, B_{M-1}\}$	possible values of output attribute B
$M(A_{i,j})$	cardinality measure of fuzzy set $A_{i,j}$
$I(B, A_i)$	cumulative joint information in attributes B and A_i

$I(B; A_i)$ cumulative mutual information between attributes B and A_i

$H(A_i)$ cumulative entropy of attribute A_i

I. INTRODUCTION

ONE of principal steps in the reliability analysis and estimation of any system is the construction of mathematical representation [1, 2]. There are some typical mathematical representations and descriptions of real systems under study in reliability engineering that include a structure function, fault tree, reliability block diagram, Markov model, and Petri Nets etc. The structure function used as a mathematical representation has been proposed. The concept of structure function is introduced in reliability engineering in order to mathematically describe a real system under study. In this case, the system is represented as a mapping that assigns a system state to every possible component states profile. Therefore, the system performance level is defined in terms of the states of all its components. It means that all possible components states and performance levels must be indicated and reflected in the structure function.

The structure function facilitates the representation of the system reliability behavior using two typical mathematical models in Reliability Engineering, i.e. Binary-State System (BSS) and Multi-State System (MSS). BSS permits only two states for a system and its components: perfect functioning and complete failure. However, in practice, many systems can exhibit different performance levels between the two extreme states of full functioning and fatal failure [1, 2]. MSS is a mathematical model that is used to describe a system with several (more than two) levels of performance [1, 3, 4]. The concept of the structure function is used to represent BSS and MSS and associates the space of component states and system performance levels. In general, the structure function is defined as $L_1 \times \dots \times L_n \rightarrow L$ (n is the number of system components). The BSS structure function is a special case if $L_1 = \dots = L_n = L = \{0, 1\}$. Therefore in this paper, we consider the MSS structure function analysis.

The analysis of system reliability using the knowledge of structure function is not new. Such attempts have been made in the classical 1975 book by Barlow and Proschan [5]. There are many methods to measure and estimate system reliability based on the structure function. Development of these methods is considered, for example, in [1, 3, 4, 6, 8]. These methods allow examining availability/reliability of MSS and

This research was supported in part by the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 610425.

E.Zaitseva and V.Levashenko are with the Department of Informatics, University of Zilina, Zilina, 01026, Slovakia (e-mail: {elena.zaitseva, vitaly.levashenko}@fri.uniza.sk).

BSS [1, 3], providing importance analysis [4, 6], calculating other reliability indices and measures [4, 7, 9]. It should be noted that structure function allows representing a system of a complex structure [6, 9]. The structure function can be defined as a discrete function [1, 10], continuous [11] or function with fuzzy states [8, 12]. Most mathematical methods used for structure function analysis are proposed for a discrete function where a limit state function must be determined. The function values are integer in this case [1, 4]. Therefore, performance level of a real system must be transformed into a set integer from 0 to $M-1$ for this function construction. A similar procedure is necessary also for the component states.

In most studies, structure function is usually assumed to be precise and ambiguities are not taken into account: the structure function defines correlation between components states and system performance level for all possible components states. Therefore, the definition and construction a structure function can be a complex problem in some cases. This means that the structure function may not be realistic in real-world applications, because data about the real system is uncertain, as a rule.

The uncertainty of initial data for the construction of the structure function can be caused by two factors. The first one is ambiguity and vagueness of collected data, because any value for this data has an inaccuracy or error of measurement. For example, this ambiguity can be caused by an error of measuring instruments. Therefore, the collected data values are associated with imprecision. The second factor is incomplete specification of data, because some values of system components states or performance levels cannot be obtained. This factor brings about some incomplete values of the system components states or performance level. However, it will be very expensive in terms of resources or time to obtain a complete set of data. Therefore, the uncertainty of initial data must be considered in the structure function construction and development of methods for its analysis.

There are two solutions for this problem. For the first one, it is necessary to develop a new reliability model that takes ambiguities and uncertainties into consideration [8, 12, 13, 14]. Uncertainties and ambiguities in a real system have been dealt with using of the possibility concept. However, it is worth pointing out that some uncertainties, which are not random in nature, may play important roles in the structure function construction [8, 13, 15]. Uncertainty can be produced due to some factors that can be evaluated solely based on an expert's experience and judgment. This uncertainty cannot be indicated in a quantitative form by probability theory. Fuzzy logic makes it possible to define the structure function in a more flexible form for such data than the probabilistic approach. Yet, application of the new mathematical model leads to a development of new mathematical methods to analyze this model.

The second solution is to use a traditional model, but involves the development of new methods to construct the structure function that takes into consideration uncertainties of the initial data. In this paper, we propose a method based on the application of Fuzzy Decision Tree (FDT). FDTs are

widely used in Data Mining for analysis of uncertain data and decision making with ambiguities. In this case, collected data for structure function construction can be defined with possibility or confidence. In addition, FDTs allow taking into account uncertainties caused by incomplete specified data. This is possible when it is expensive to obtain all data about real system behavior or there is little data with poor documentation. As a rule, if the exact value of the actual data about the system behavior cannot be determined, we need to rely on more data to give additional information necessary to correct the theoretical model used [13, 15]. An FDT allows reconstructing these data with different levels of the possibility (confidence) [16, 17].

The use of FDTs for construction of the structure function assumes induction of a tree based on the data (fuzzy and/or crisp) about the real system behavior and these data can be incompletely specified. Structure function values are then defined for all combinations of component states by the FDT: component states are interpreted as FDT attributes and the structure function value is aligned with one of the M values (classes) for the system performance level.

This paper is structured as follows. Section 2 discusses the concept of structure function. Principal steps of the proposed method are considered in sections 3 – 5:

- collection of data in the repository (section 3);
- representation of the system model in the form of an FDT (section 4);
- construction of the structure function based on the FDT (section 5).

The detailed process of FDT induction is presented in section 4, including basic rules and mathematical background of this process. We use measures of cumulative information estimates for FDT construction [18]. These measures are defined based on possibilistic approach. So, collected data should satisfy the assumptions of this approach. At the same time, for FDT induction, this data is interpreted as fuzzy data. According to paper [19], the probability data can be interpreted as fuzzy data if the sum of its membership degree equals 1.

The proposed method is illustrated by a manual calculation of the example for detailed representation and understanding of the concept and principal steps of the method. This example is continued in section 5 to explain the construction of structure function based on FDT. In addition, the investigation of accuracy of proposed method for other systems is shown in this section too.

II. BACKGROUND FOR THE NEW METHOD OF STRUCTURE FUNCTION CONSTRUCTION

A. Structure function of the system

A system can be represented in several ways in terms of reliability engineering [10]. It can be a fault tree, reliability block diagram or structure function. The structure function is the most general representation form which captures the relationships between the components of a system and the system itself in such a way that the state of the system is

elicited from the states of its components through the structure function.

Consider the MSS structure function as a more general mathematical model. Suppose the system can be divided into n components (subsystems). State of each component can be denoted by a random variable, x_i , that takes on the value $x_i = 0$ if the component fails in a stationary state and $x_i = 1, \dots, m_{i-1}$ if the component is functioning. The system has M performance levels.

Denote $\phi(\mathbf{x})$ as the structure function, then:

$$\phi(\mathbf{x}) = \phi(x_1, \dots, x_n): \{0, \dots, m_1 - 1\} \times \dots \times \{0, \dots, m_n - 1\} \rightarrow \{0, \dots, M - 1\}, \quad (1)$$

where $\phi(\mathbf{x})$ is the system state (performance level) from failure ($\phi(\mathbf{x}) = 0$) to perfect functioning ($\phi(\mathbf{x}) = M - 1$); $\mathbf{x} = (x_1, \dots, x_n)$ is the state vector; x_i is the i -th component state that changes from failure ($x_i = 0$) to perfect functioning ($x_i = m_i - 1$).

Note that the structure function of MSS (1) is transformed into the structure function of BSS if $m_i = M = 2$ (for all $i \in \{1, \dots, n\}$):

$$\phi(\mathbf{x}) = \phi(x_1, \dots, x_n): \{0, 1\}^n \rightarrow \{0, 1\} \quad (2)$$

In this paper, we consider coherent systems assuming that: (a) the improvement of any component does not degrade the state of the system; that is the system structure function is monotonic: $\phi(x_i, \mathbf{x}) \leq \phi(x_j, \mathbf{x})$ for any $x_i \leq x_j$; and (b) there is no irrelevant component in the system.

Many reliability indices and measures can be calculated based on the system structure function. One of them is the probability of the system performance level that is calculated as follows [4, 6]:

$$A_j = \Pr \{ \phi(\mathbf{x}) = j \}, \quad (3)$$

where every system component is characterized by the probabilities of its state:

$$p_{i,s} = \Pr \{ x_i = s \}, s = 0, \dots, m_i - 1. \quad (4)$$

For example, consider the simple series-parallel system in Fig. 1 of three components ($n = 3$). This system has three performance levels ($M = 3$): 0 – non-operational, 1 – partially operational, 2 – fully operational. Two components (x_1 and x_2) have only 2 possible states ($m_1 = m_2 = 2$): functional (state 1) or dysfunctional (state 0). The third component has 3 quality levels ($m_3 = 3$): from 0 (it is faulted) to 2 (it is perfectly functioning). The structure function of this system is defined in Table 1. The probabilities of this system performance level (3) are shown in Table 2. These probabilities are calculated as sum of probabilities of states (state vectors) that correspond with fixed performance levels. Formally, the structure function arranges all possible states into M mutually exclusive classes that comply with the system performance levels.

The structure function facilitates the calculation of the boundary system states [18], minimal cut/path sets [7] and importance measures [18, 21]. However, defining structure function (1) for a real application can be difficult.

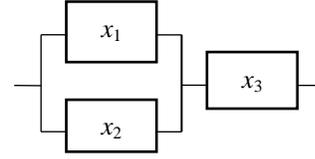


Fig. 1. The series-parallel system

TABLE I
STRUCTURE FUNCTION OF A SIMPLE SERVICE SYSTEM

Component states		x_3		
x_1	x_2	0	1	2
0	0	0	0	0
0	1	0	1	1
1	0	0	1	1
1	1	0	2	2

TABLE II
PROBABILITIES OF THE SYSTEM PERFORMANCE LEVELS

System performance levels	State vectors	Probabilities of the system performance levels
0	(0 0 0), (0 0 1), (0 1 0), (0 1 1), (0 0 1), (1 0 2)	$p_{3,0} + p_{1,0}p_{2,0}(p_{3,1} + p_{3,2})$
1	(0 1 1), (1 0 1), (0 1 2), (1 0 2)	$(p_{1,0}p_{2,1} + p_{1,1}p_{2,0})(p_{3,1} + p_{3,2})$
2	(1 1 1), (1 1 2)	$p_{1,1}p_{2,1}(p_{3,1} + p_{3,2})$

B. General description of the new method for structure function construction

As a rule, structure function can be defined by the system structure analysis or based on expert data [22, 23, 24]. In system structure analysis, the system is interpreted as a set of components (a subsystem) with correlations. These correlations can be defined by functional relations that are interpreted as the structure function (1). An example of such a system is one with a typical structure such as series-parallel (see Fig.1), bridge or k -out-of- n systems. Still, there are many structure-complex systems for which correlations and/or connections of components are latent or uncertain (for example power systems, network systems). The construction of the structure function for this system is complex. Other representations are used and special methods are developed in reliability estimation for such systems [11, 25]. Expert data for description and representation of the system can be used in this case.

The construction of a structure function based on expert data requires special analysis and transformation of initial data [26, 27], because this data is uncertain. The uncertainty can be caused by a lot of factors, but we have only considered two of them. The first factor is ambiguity and vagueness of collected data values. This type of ambiguity can be caused by an inaccuracy or error of measurement, expert subjective evaluation etc. For example, two experts can set different values of system performance level for an equal situation [28, 29] or distance measurement equipment can have an error [27, 30]. The second factor is incomplete specification of data, because some values of system components states or performance levels cannot be obtained. This factor is caused

by impossibility to indicate some values of the system components states or performance level, because it can be very expensive or incurs long unacceptable time.

Therefore, the structure function construction must address two aspects. The first is structure function definition as a mapping assigning the system performance level to each possible profile of component states (for example see Table 2). The second is uncertainty of initial data that is caused by ambiguity of data values and incomplete specification of all possible profiles of components states. In other words, this problem can be interpreted as a classification problem for uncertain data. It is a typical problem of Data Mining and one of possible decision is the application of Decision Tree or Fuzzy Decision Tree (FDT) [31-34].

We suggest a new method for the construction of a structure function (1) that is based on the application of the FDT [18, 34, 35]. This method includes the following step (Fig.2):

- Collection of data in the repository according to requests of FDT induction;
- Representation of the system model in the form of an FDT that classifies components states according to the system performance levels;
- Construction of the structure function as decision table that is created by inducted FDT.

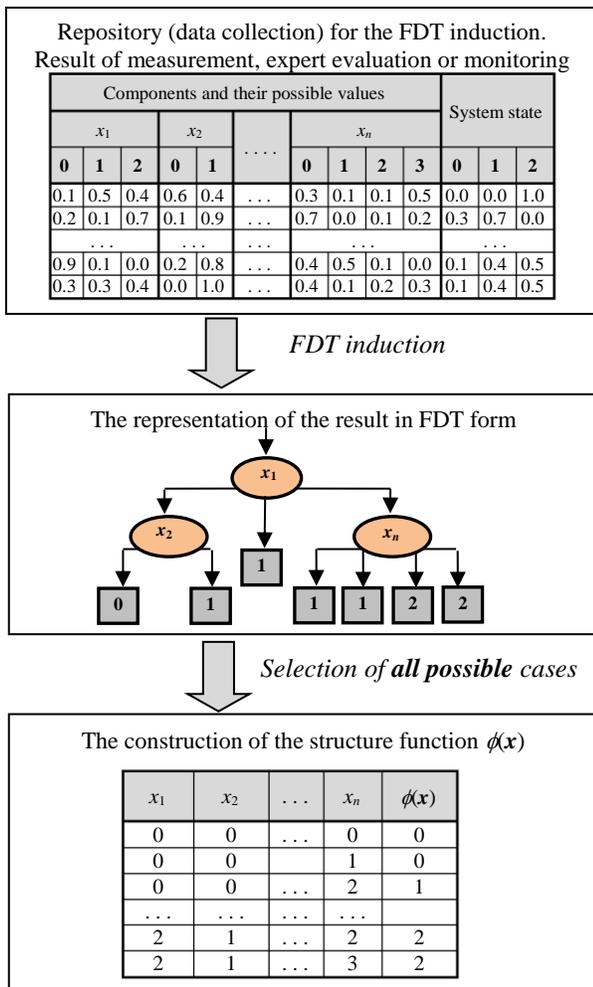


Fig.2 Principal steps for the structure function construction

Therefore, the structure function is constructed as a decision table that classifies the system performance level based on each possible profile of components states. The decision table is formed based on FDT that provides the mapping for all possible components states (input data) in M performance levels. FDT is inducted by uncertain data that is presented in form of specified repository.

III. COLLECTION OF DATA IN THE REPOSITORY

Collection of data in the form of a repository is provided by the monitoring or expert evaluation of values of system component states and system performance level. The repository for the FDT induction is presented as table (Fig.2). The columns number is $n+1$ (for indication of n components and the system performance level). Every n column is separated into m_i sub-columns and the column for the system performance level has M sub-columns. The sub-column is assigned with one of the values of component states or performance levels. Every row of the table represents one monitoring situation or evaluation. The table cell includes number (from 0 to 1) that is interpreted as the possibility of this value. Note that the sum of these possibilities for each value equals to 1. Such data can be obtained from expert evaluations or possibilistic fuzzy clustering [36, 37]. These possibilities correspond to a membership function of fuzzy data [19]. This demand for initial data representation is caused by the method of FDT induction. Therefore, values of the i -th component state and the system performance levels are defined by possibilities. These possibilities indicate ambiguity of collected data values for the analysis. Having indicated and considered the uncertainty of the monitoring data, it is possible to increase the accuracy of the result from this data analysis.

For example, consider a simple service system (Fig. 3). The system consists of three components ($n = 3$): service point 1 (x_1), service point 2 (x_2) and infrastructure (x_3). This system has three performance levels ($M = 3$): 0 – non-operational (no customer is satisfied), 1 – partially operational (some customers are satisfied), 2 – fully operational (all customers are satisfied). The service points have only two possible states ($m_1 = m_2 = 2$): functional (state 1) or dysfunctional (state 0). The infrastructure can be modelled by 4 quality levels ($m_3 = 4$), i.e. from 0 (the quality of the infrastructure is poor) to 3 (the quality is perfect).

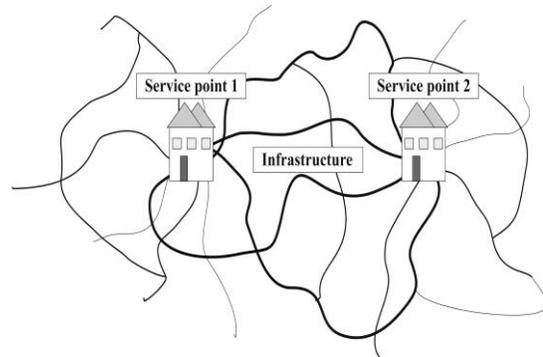


Fig. 3. A simple service system

The monitoring of this system results in information obtained about the system performance level and appropriate component states. However, this information is uncertain because the data from real monitoring is not complete and values are ambiguous. These data can be defined with a possibility ranging from 0 to 1. For example, the monitoring of the service system indicates the state of the service point 1 as functional ($x_1 = 1$) with the possibility of 0.1 and dysfunctional ($x_1 = 0$) with the possibility of 0.9. Result of monitoring the simple service system shown in Fig.3 is represented in Table 3.

Let us illustrate the correlation of monitoring data and the structure function for this system. The monitored data can be transformed into a structure function (1) for the simple service system based on this rule: only the value with the highest possibility is considered. For example, the variable x_1 in Table 3 has value 0 with the possibility of 0.9 and value 1 with the possibility of 0.1. The resultant value is defined as 0 in this case. The transformation of the monitored data for the simple service system in Table 3 is presented in the Table 4.

It can be seen that the data in Table 4 cannot be interpreted as a structure function (1) because there are some component states (state vectors) that are not indicated by the monitoring. The value of the system performance level for state vectors $\mathbf{x} = (0\ 0\ 0)$, $\mathbf{x} = (1\ 0\ 3)$, $\mathbf{x} = (1\ 1\ 1)$ and $\mathbf{x} = (1\ 1\ 2)$ has not been obtained by this monitoring. Traditional mathematical approach for system reliability analysis based on the structure function cannot be used in this case. Therefore, construction of a structure function (1) based on incomplete monitored data requires a special transformation and development of new methods. In this paper, we suggest a new method for the construction of a structure function based on FDT. This method allows the reduction of indeterminate values and to obtain a fully specified structure function.

IV. REPRESENTATION OF SYSTEM MODEL IN THE FORM OF AN FDT

Representation of system model in the form of an FDT is determined by the correlation between system performance level and system component states based on uncertain data [32, 35].

A decision tree (and FDT in particular) can be considered as an alternative form of the structure function. The structure function maps states vector to each equivalence class of system performance levels. At the same time, a decision tree is a formalism for expressing mappings of input attributes (components states) and output attribute/attributes (system performance level), consisting of an analysis of attribute nodes linked to two or more sub-trees and leaves or decision nodes labeled with a class (in our case it is the system performance level) [31]. Analysis of a node produces some outcome based on attribute values of an instance, where each possible outcome is associated with one of the sub-trees. An instance is classified by the starting point at the root node of the tree. If this node is not a leaf, the outcome for the instance is determined and the process continues using the appropriate sub-tree. When a leaf is eventually encountered, its label gives the predicted class of the instance. The system component states are interpreted as values of the input attributes. The system performance levels are considered as an instance that is classified into M class.

FDT is one of the possible types of decision trees that operate with fuzzy data (attributes) and methods of fuzzy logic. Construction of a structure function assumes operation with ambiguous data and the analysis of these data can be implemented based on the methods of fuzzy logic that is caused by ambiguous of data values [8, 12, 16, 32]. The ambiguity may be present in obtaining numeric values of the attributes (system components states) and in obtaining the exact class where the instance (system performance level) belongs to.

TABLE III
MONITORING DATA OF THE SERVICE SYSTEM

No.	x_1		x_2		x_3			$\phi(\mathbf{x})$			
	0	1	0	1	0	1	2	3	0	1	2
1	0.9	0.1	0.8	0.2	0.3	0.7	0.0	0.0	0.9	0.1	0.0
2	0.8	0.2	0.9	0.1	0.0	0.1	0.8	0.1	0.7	0.3	0.0
3	1.0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	0.6	0.4	0.0
4	0.9	0.1	0.1	0.9	1.0	0.0	0.0	0.0	1.0	0.0	0.0
5	1.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0
6	0.8	0.2	0.1	0.9	0.0	0.1	0.8	0.1	0.0	1.0	0.0
7	0.1	0.9	1.0	0.0	1.0	0.0	0.0	0.0	0.9	0.1	0.0
8	0.0	1.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0
9	0.0	1.0	1.0	0.0	0.0	0.1	0.9	0.0	0.2	0.8	0.0
10	1.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0.9	0.1
11	0.1	0.9	0.1	0.9	1.0	0.0	0.0	0.0	0.9	0.1	0.0
12	0.0	1.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0

TABLE IV
MONITORING DATA OF THE SERVICE SYSTEM

Component states	x_1	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1
	x_2	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1
	x_3	0	1	2	3	0	1	2	3	0	1	2	3	0	1	2
System performance level	$\phi(\mathbf{x})$	-	0	0	0	0	1	1	1	0	1	1	-	0	-	2

A. Fuzzy Decision Trees

There are different methods to induce an FDT [16, 18, 32-34]. The principal goal of these methods for FDT induction is to select expanded attributes and determine the leaf node. The key point of the methods used for induction of FDT is a heuristic for selecting expanded attributes. The FDT induction is implemented based on some initial data that are interpreted as a training test. As a rule, these data are obtained by monitoring and consist of a set input attributes and output attributes.

The FDT induction is implemented by the definition of the correlation between input n attributes A_1, \dots, A_n and an output attribute B . The construction of the system structure function supposes that the system performance level is the output attribute and component states (state vectors) are input attributes. Each input attribute (component states) A_i ($1 \leq i \leq n$) is measured by a group of discrete values from 0 to m_i-1 that concur with the values of the i -th component states: $\{A_{i,0}, \dots, A_{i,j}, \dots, A_{i,m_i-1}\}$. The FDT assumes that the input set A_1, \dots, A_n is classified as one of the output value attributes B . The output attribute value B_w agrees with one of the system performance levels and is defined as M values ranging from 0 to $M-1$ ($w = 0, \dots, M-1$).

For example, a set of input attributes $\{A_1, A_2, A_3\}$ and output attribute B for the service system in Fig. 2 are indicated in Table 5 according to FDT terminology. Each attribute is defined as: $A_1 = \{A_{1,0}, A_{1,1}\}$, $A_2 = \{A_{2,0}, A_{2,1}\}$, $A_3 = \{A_{3,0}, A_{3,1}, A_{3,2}, A_{3,3}\}$ and $B = \{B_0, B_1, B_2\}$.

TABLE V
ATTRIBUTE VALUES

Attribute	Attribute Values	Attribute Value Description
A ₁	A _{1,0}	The first service point is dysfunctional
	A _{1,1}	The first service point is functional
A ₂	A _{2,0}	The second service point is dysfunctional
	A _{2,1}	The second service point is functional
A ₃	A _{3,0}	The quality of the infrastructure is poor
	A _{3,1}	The quality of the infrastructure is sufficient
	A _{3,2}	The quality of the infrastructure is good
	A _{3,3}	The quality of the infrastructure is perfect
B	B ₀	The system is non-operational (no customer is satisfied)
	B ₁	The system is partially operational (some customers are satisfied)
	B ₂	The system is fully operational (all customers are satisfied)

A fuzzy set A with respect to a universe U is characterized by a membership function $\mu_A : U \rightarrow [0,1]$, assign a A -membership degree, $\mu_A(u)$, to each element u in U . $\mu_A(u)$ gives us an estimation of u belonging to A . The cardinality measure of the fuzzy set A is defined by $M(A) = \sum_{u \in U} \mu_A(u)$, which is the measure of the size of A .

For $u \in U$, $\mu_A(u)=1$ means that u is definitely a member of A and $\mu_A(u)=0$ means that u is definitely not a member of A , while $0 < \mu_A(u) < 1$ means that u is partially a member of A . If either $\mu_A(u)=0$ or $\mu_A(u)=1$ for all $u \in U$, A is a crisp set. The set of input attributes A is crisp for which $\mu_A(u)=0$ or $\mu_A(u)=1$. The values of input attributes and output attribute are defined by the membership function in Table 6 for the service system in Fig. 3. These values are obtained based on the monitored data in Table 3 and are used for the FDT construction as a training test. The cardinality measure as the sum of output attribute values is in the last row in Table 6.

In this paper, the method for FDT induction is used for the construction of the structure function. This method based on cumulative information estimates [18, 34]. The cumulative information estimates facilitates the definition of the criterion of expanded attribute selection to induct FDT with different properties [38]. These estimates are calculated by measures of entropy and information. The measures of entropy and information have been introduced in information theory that is based on probabilistic approach. The correct application of these measures demands that the sum of possibilities of all values of every attribute equals 1 [19, 36, 37]. Note, the possibility of attribute's value in terms of FDT induction is measured as confidence degree or degree of truth in this value.

B. Fuzzy Decision Trees Induction

To induce an FDT, the method based on the cumulative information estimates [18] is used. These estimates facilitate the induction of an FDT with different properties. Criteria for building non-ordered, ordered or stable FDTs have been considered in [34, 35]. This selection criterion is defined as a different type of cumulative mutual information $I(B; A)$, where B and A are output and input attributes (or their values).

Our method is illustrated as a non-ordered FDT in this paper. The selection criterion of expanded attributes A_{i_j} for induction of a non-ordered FDT is defined as:

TABLE VI
A TRAINING SET FOR THE FDT INDUCTION

No.	A ₁		A ₂		A ₃				B			
	A _{1,0}	A _{1,1}	A _{2,0}	A _{2,1}	A _{3,0}	A _{3,1}	A _{3,2}	A _{3,3}	A _{1,0}	A _{1,1}	A _{2,0}	
1	0.9	0.1	0.8	0.2	0.3	0.7	0.0	0.0	0.9	0.1	0.0	
2	0.8	0.2	0.9	0.1	0.0	0.1	0.8	0.1	0.7	0.3	0.0	
3	1.0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	0.6	0.4	0.0	
4	0.9	0.1	0.1	0.9	1.0	0.0	0.0	0.0	1.0	0.0	0.0	
5	1.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	
6	0.8	0.2	0.1	0.9	0.0	0.1	0.8	0.1	0.0	1.0	0.0	
7	0.1	0.9	1.0	0.0	1.0	0.0	0.0	0.0	0.9	0.1	0.0	
8	0.0	1.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	
9	0.0	1.0	1.0	0.0	0.0	0.1	0.9	0.0	0.2	0.8	0.0	
10	1.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0.9	0.1	
11	0.1	0.9	0.1	0.9	1.0	0.0	0.0	0.0	0.9	0.1	0.0	
12	0.0	1.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	
M(B)/12										0.433	0.475	0.092

$$i_q = \operatorname{argmax} \left(\frac{\mathbf{I}(\mathbf{B}; \mathbf{A}_{i_1, j_1}, \dots, \mathbf{A}_{i_{q-1}, j_{q-1}}, \mathbf{A}_{i_q})}{\operatorname{Cost}(\mathbf{A}_{i_q}) \times \mathbf{H}(\mathbf{A}_{i_q})} \right), \quad (5)$$

where $\mathbf{A}_{i_1, j_1}, \dots, \mathbf{A}_{i_{q-1}, j_{q-1}}$ are values of input attributes $\mathbf{A}_{i_1}, \dots, \mathbf{A}_{i_{q-1}}$ of the path from the root node to the examined attribute; \mathbf{A}_{i_q} is the attribute that is not in this path; $\mathbf{I}(\mathbf{B}; \mathbf{A}_{i_1, j_1}, \dots, \mathbf{A}_{i_{q-1}, j_{q-1}}, \mathbf{A}_{i_q})$ is cumulative mutual information. $\operatorname{Cost}(\mathbf{A}_{i_q})$ is an integrated measure that covers financial and temporal costs required to define the value of the \mathbf{A}_{i_q} for an instance and this value is defined a priori. $\mathbf{H}(\mathbf{A}_{i_q})$ is a cumulative entropy of this input attribute \mathbf{A}_{i_q} .

The cumulative mutual information in output attribute B about the attribute \mathbf{A}_{i_q} and the sequence of values $U_{q-1} = \{\mathbf{A}_{i_1, j_1}, \dots, \mathbf{A}_{i_{q-1}, j_{q-1}}\}$ reflects the influence of attribute \mathbf{A}_{i_q} on the output attribute B when sequence U_{q-1} is known. This measure has been introduced in [18] and calculated as:

$$\mathbf{I}(\mathbf{B}; U_{q-1}, \mathbf{A}_{i_q}) = \sum_{j_q=0}^{m_q-1} \sum_{j=0}^{m_b-1} \mathbf{M}(\mathbf{B}_j \times \mathbf{A}_{i_1, j_1} \times \dots \times \mathbf{A}_{i_q, j_q}) \times \left(\mathbf{I}(\mathbf{B}_j, U_{q-1}) + \mathbf{I}(U_{q-1}, \mathbf{A}_{i_q, j_q}) - \mathbf{I}(\mathbf{B}_j, U_{q-1}, \mathbf{A}_{i_q, j_q}) - \mathbf{I}(U_{q-1}) \right) \quad \text{bits}, \quad (6)$$

where $\mathbf{M}(\mathbf{B}_j \times \mathbf{A}_{i_1, j_1} \times \dots \times \mathbf{A}_{i_q, j_q})$ is a cardinality measure of fuzzy set $\mathbf{B}_j \times \mathbf{A}_{i_1, j_1} \times \dots \times \mathbf{A}_{i_q, j_q}$; $\mathbf{I}(\mathbf{B}_j, U_{q-1})$, $\mathbf{I}(U_{q-1}, \mathbf{A}_{i_q, j_q})$, $\mathbf{I}(\mathbf{B}_j, U_{q-1}, \mathbf{A}_{i_q, j_q})$ and $\mathbf{I}(U_{q-1})$ are cumulative joint information.

Note, that this cumulative mutual information equals the difference between cumulative condition entropies:

$$\mathbf{I}(\mathbf{B}; U_{q-1}, \mathbf{A}_{i_q}) = \mathbf{H}(\mathbf{B} | U_{q-1}) - \mathbf{H}(\mathbf{B} | U_{q-1}, \mathbf{A}_{i_q}).$$

The entropy $\mathbf{H}(\mathbf{B} | U_{q-1})$ describes the uncertainty of a situation if values of previous attributes U_{q-1} are known and value of attribute \mathbf{A}_{i_q} is unknown. The next cumulative condition entropy $\mathbf{H}(\mathbf{B} | U_{q-1}, \mathbf{A}_{i_q})$ describes a situation when values of attributes U_{q-1} and \mathbf{A}_{i_q} are known.

Let us have a sequence of $q-1$ input attributes $\mathbf{A}_{i_1}, \dots, \mathbf{A}_{i_{q-1}}$ and one output attribute B. The cumulative joint information of the sequence of values $U_{q-1} = \{\mathbf{A}_{i_1, j_1}, \dots, \mathbf{A}_{i_{q-1}, j_{q-1}}\}$ ($q \geq 2$) and value \mathbf{B}_j is:

$$\mathbf{I}(\mathbf{B}_j, U_{q-1}) = \log_2 N - \log_2 \mathbf{M}(\mathbf{B}_j \times \mathbf{A}_{i_1, j_1} \times \dots \times \mathbf{A}_{i_{q-1}, j_{q-1}}) \quad \text{bits}. \quad (7)$$

The cumulative entropy of input attribute \mathbf{A}_{i_q} reflects the ambiguity of this attribute. We calculate this entropy using the next rule:

$$\mathbf{H}(\mathbf{A}_{i_q}) = \sum_{j_q=0}^{m_q-1} \mathbf{M}(\mathbf{A}_{i_q, j_q}) \times \mathbf{I}(\mathbf{A}_{i_q, j_q}) \quad \text{bits}, \quad (8)$$

where $\mathbf{M}(\mathbf{A}_{i_q, j_q})$ is the cardinality measure of set \mathbf{A}_{i_q, j_q} , and $\mathbf{I}(\mathbf{A}_{i_q, j_q})$ is the cumulative joint information.

Maximum value i_q in (5) facilitates the selection of expanded attribute \mathbf{A}_{i_q} . This attribute will be associated with a node of the FDT.

There are two tuning thresholds α and β in this method of FDT induction [18, 32]. A tree branch stops to expand when either the frequency f of the branch is below α or when more than β percent of instances left in the branch has the same class label. These values are thus key parameters needed to decide whether we have already arrived at a leaf node or whether the branch should be expanded further. Decreasing the parameter α and increasing the parameter β allow us to build large FDTs. On one hand, large FDTs describe datasets in more detail. On the other hand, these FDTs are very sensitive to noise in the dataset. We empirically select parameters $\alpha = 0.10$ and $\beta = 0.90$. We estimate that a confidence degree of more than 0.90 would allow us to reach a decision with sufficient confidence. Moreover, the threshold frequency 0.10 eliminates the variants of no-principal decisions. Notably, increasing the size of the FDT has no influence on the FDT root or the higher FDT nodes. It only adds new nodes and leaves to the bottom part of the FDT. These new nodes and leaves have a low bearing on decision making.

A learning algorithm for construction non-ordered fuzzy decision trees can be described as follows [18].

Input data: The training dataset (as example, see Table 6).

$$\operatorname{Attr} = \{\mathbf{A}_1, \dots, \mathbf{A}_n\}; \quad q=0; \quad U_q = \emptyset$$

Output data: non-ordered Fuzzy Decision Tree.

Tree = **buildTree** (U_q, Attr)

1. Calculate cumulative mutual information (6)
2. Select attribute with the maximal value of criterion (5)

$$i_q = \operatorname{argmax} \mathbf{I}(\mathbf{B}; U_q, \mathbf{A}_{i_q}) / \operatorname{Cost}(\mathbf{A}_{i_q}) \text{ for } \forall \mathbf{A}_{i_q} \in \operatorname{Attr}$$
3. Assign current node

$$\text{Tree} \leftarrow \text{node}(\mathbf{A}_{i_q}); \quad \operatorname{Attr} = \operatorname{Attr} \setminus \mathbf{A}_{i_q}$$
4. Choose leaves and continue

$$\text{for } (\forall \mathbf{A}_{i_q, j}, j = 0, \dots, m_q - 1)$$
 - { $q++$
 - $U_q = U_{q-1} \cup \mathbf{A}_{i_q, j}$
 - if ($\mathbf{A}_{i_q, j}$ is leaf) $\text{Tree} \leftarrow \text{leaf}(\mathbf{A}_{i_q, j})$
 - else Recursively construct the sub-trees:

$$\text{Tree} = \mathbf{buildTree}(U_q, \operatorname{Attr}) \}$$

Let us explain the technique of computations by the example below. Continuing with the monitored data from the example of the simple service system, let us build a non-ordered FDT with parameters $\beta=0.90$ and $\alpha=0.10$ (see data from Table 6). We will allow that values of integrated measure $\operatorname{Cost}(\mathbf{A}_i)$ of each attribute are equal for simplification of calculation: $\operatorname{Cost}(\mathbf{A}_i)=1$ for each $i = 1, \dots, 4$.

Let us show the cumulative information estimates based on one input attribute A_3 only. The attribute A_3 has four possible values: $A_{3,0}$, $A_{3,1}$, $A_{3,2}$ and $A_{3,3}$. Similarly, the output attribute B has three possible values: B_0 , B_1 , and B_2 .

The cumulative mutual information (6) in output attribute B about attribute A_3 equals:

$$\begin{aligned} \mathbf{I}(B; A_3) &= \sum_{j_q=0}^3 \sum_{j=0}^2 M(B_j \times A_{3,j_q}) \times \left(\mathbf{I}(B_j) + \mathbf{I}(A_{3,j_q}) - \mathbf{I}(B_j, A_{3,j_q}) \right) = \\ &= \sum_{j=0}^2 M(B_j) \times \mathbf{I}(B_j) + \sum_{j_q=0}^3 M(A_{3,j_q}) \times \mathbf{I}(A_{3,j_q}) - \sum_{j_q=0}^3 \sum_{j=0}^2 M(B_j \times A_{3,j_q}) \times \mathbf{I}(B_j, A_{3,j_q}) = \\ &= \mathbf{H}(B) + \mathbf{H}(A_3) - \mathbf{H}(B, A_3), \end{aligned}$$

where $\mathbf{H}(B)$, $\mathbf{H}(A_3)$ and $\mathbf{H}(B, A_3)$ are the cumulative entropies.

We have to calculate the cumulative entropies based on equations (8) and (7):

$$\begin{aligned} \mathbf{H}(B) &= N \times \log_2(N) - \sum_{j=0}^2 M(B_j) \times \log_2(B_j) = 12 \times \log_2 12 - \\ &- 5.2 \times \log_2 5.2 - 5.7 \times \log_2 5.7 - 1.1 \times \log_2 1.1 = 16.19 \text{ bit.} \end{aligned}$$

Similarly,

$$\begin{aligned} \mathbf{H}(A_3) &= 12 \times \log_2 12 - 3.3 \times \log_2 3.3 - 3.0 \times \log_2 3.0 - \\ &- 2.5 \times \log_2 2.5 - 3.2 \times \log_2 3.2 = 23.91 \text{ bit.} \end{aligned}$$

$$\begin{aligned} \mathbf{H}(B, A_3) &= \mathbf{H}(B_0, A_{3,0}) + \mathbf{H}(B_0, A_{3,1}) + \mathbf{H}(B_0, A_{3,2}) + \\ &+ \mathbf{H}(B_0, A_{3,3}) + \mathbf{H}(B_1, A_{3,0}) + \dots + \mathbf{H}(B_2, A_{3,3}) = 34.55 \text{ bit.} \end{aligned}$$

Therefore, the cumulative mutual information equals:

$$\begin{aligned} \mathbf{I}(B; A_3) &= \mathbf{H}(B) + \mathbf{H}(A_3) - \mathbf{H}(B, A_3) = \\ &= 16.19 + 23.91 - 34.55 = 5.54 \text{ bit.} \end{aligned}$$

We calculate all of the values $\mathbf{I}(B; A_i) / \mathbf{H}(A_i)$ for each i ($i = 1, \dots, 4$). The minimal value 5.54/23.91 corresponds to the input attribute A_3 . Therefore, we assign this attribute A_3 as a FDT root node. The first level of the FDT inducted according to (4-7) for the dataset in Table 6 is shown in Fig. 4.

Let us explain the first level of the FDT in more detail.

Preliminary analysis of the dataset (see Table 6) shows that possible values of the output attribute B are calculated as $M(B_j)/N$. Therefore, the values of attribute B distributed as follows: value 0 – with confidence $M(B_0)/12 = 5.2/12 = 0.433$,

value 1 – with confidence $M(B_1)/12 = 5.7/12 = 0.475$ and value 2 – with confidence $M(B_2)/12 = 1.1/12 = 0.092$ only.

The attribute A_3 has the maximum value of criterion (5). Therefore this attribute is associated with the FDT root (top node). This attribute can have the following possible values: $A_{3,0}$, $A_{3,1}$, $A_{3,2}$ and $A_{3,3}$. These values associated with branches of FDT. The frequencies of each j branch of the root node are calculated as $M(A_{q,j})/N$. For our examples, the frequency of branch $A_{3,0}$ equals $f(A_{3,0}) = 3.3/12 = 0.275$. Similarly, $f(A_{3,1}) = 3.0/12 = 0.25$; $f(A_{3,2}) = 2.5/12 = 0.208$ and $f(A_{3,3}) = 3.2/12 = 0.267$. These frequencies of each branch are higher than the given threshold $\alpha=0.10$. Therefore, the data for these branches are analyzed further. The value of frequencies for each of the branch U_q on next levels of FDT are calculated as $M(A_{q,j} \times U_{q-1})/N$.

Let us calculate the values of the output attributes for each of the leaves on first level of FDT. These values for branch $U_q = \{A_3\}$ are calculated as $M(B_j \times U_{q,j})/M(U_q)$. Therefore, the frequencies of the output attribute for branch $A_{3,0}$ are $M(B_0 \times A_{3,0}) = 3.07/3.3 = 0.930$, $M(B_1 \times A_{3,0}) = 0.23/3.3 = 0.070$ and $M(B_2 \times A_{3,0}) = 0.0$. So, the value of attribute $A_{3,0}$ makes the output attribute B to be B_0 (the system is non-operational) with the confidence of 0.93. The other variants B_1 (the system is partially operational) and B_2 (the system is fully operational) of the output attribute B can be chosen with the confidence of 0.07 and 0.00. Confidence of 0.93 for the value of B_0 is more than $\beta=0.90$. Therefore, we stop the process of constructing the FDT for this branch.

If the attribute A_3 has other values $A_{3,1}$, $A_{3,2}$ or $A_{3,3}$ then value B_1 (the system is partially operational) of attribute B should be chosen with the confidences of 0.760, 0.704 and 0.447 respectively. These confidences are less than the priority threshold of the output attribute ($\beta=0.90$). Therefore, we have to continue the process of constructing other levels of FDT for these branches.

Also, we can analyze the information transformation process. Initial cumulative entropy $\mathbf{H}(B)$ equals 16.188 bit. We have obtained $\mathbf{I}(B; A_3)=5.540$ bit of information when value of attribute A_3 becomes known. Unknown information is described by cumulative conditional entropy $\mathbf{H}(B|A_3) = 10.648$. This entropy $\mathbf{H}(B|A_3)$ splits into four branches $\mathbf{H}(B|A_{3,0}) = 1.204$ bit, ..., $\mathbf{H}(B|A_{3,3})=4.868$ bit.

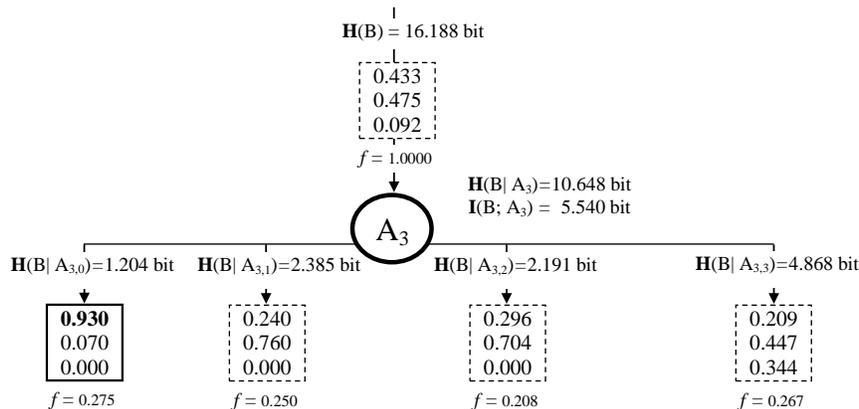


Fig. 4. The first level of the FDT inducted for the monitoring data (Table 3)

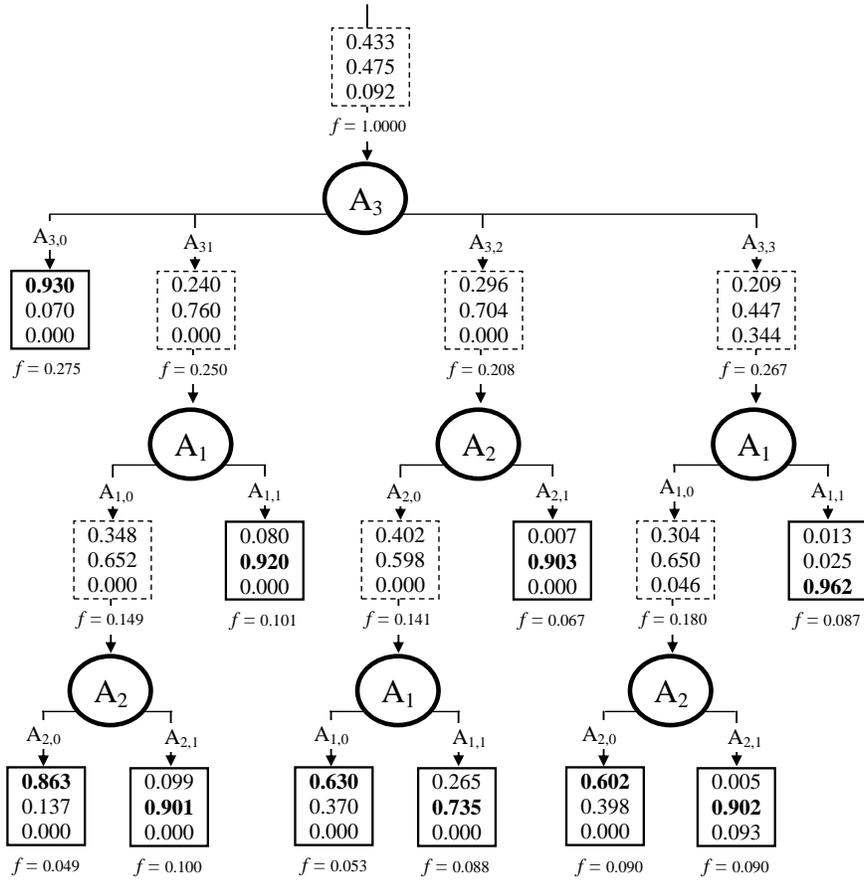


Fig. 5. Non ordered FDT induced for the monitoring data from Table 3.

Similarly, the process of inducing a non-ordered FDT is continued for the second and the third level of the FDT. The final version of the non-ordered FDT is shown at Fig.5.

V. CONSTRUCTION OF THE STRUCTURE FUNCTION BASED ON THE FDT

A. Example of construction of the structure function based on the FDT

According to [34, 35], FDTs will result in fuzzy decision rules or a decision table. A decision table indicates all possible values of input attributes and a corresponding value of the output attribute that is calculated according to the FDT. Interpretation of FDT terminology for the problem of the structure function construction is shown in Table 7. In the case of structure function construction (see Table 7), the input attributes A_i agree with the system components x_i ($i = 1, \dots, n$) and the output attribute B is assigned with the system performance level $\phi(x)$. Therefore, calculation of the decision table for all possible values of the components states is determined as the structure function of the system according to (1).

For example, construct a structure function of the simple service system (Fig.3) based on the FDT (Fig.5) that was induced based on the monitored data from Table 3. The structure function of this system corresponds to the decision table for this FDT. Therefore, all possible values of the component states (all vector states) must be analyzed by the

FDT, implying classification of the vector states into M classes of the system performance levels.

TABLE VII
CORRELATION OF THE TERMINOLOGIES OF FDT AND RELIABILITY ANALYSIS

FDT	System reliability
Number of input attribute: n	Number of the system components: n
Attribute A_i ($i = 1, \dots, n$)	System component x_i ($i = 1, \dots, n$)
Attribute A_i values: $\{A_{i,0}, \dots, A_{i,j}, \dots, A_{i,m_i-1}\}$	The i -th system component state: $\{0, \dots, m_i-1\}$
Output attribute B	System performance level $\phi(x)$
Values of output attribute B: $\{B_0, \dots, B_{M-1}\}$	System performance level values: $\{0, \dots, M-1\}$
Decision table	Structure function

Each non-leaf node is associated with an attribute $A_i \in A$, or in terms of reliability analysis: each non-leaf node is associated with a component. The non-leaf node of the attribute A_i has m_i outgoing branches. The s -th outgoing branch ($s = 0, \dots, m_i-1$) from the non-leaf node A_i agrees with the value s of the i -th component ($x_i = s$). The path from the top node to the leaf indicates the vector state of the structure function by the values of attributes and the value of the output attribute corresponds to the system performance level. If any attribute is absent in the path, then all possible values of the states are defined for the associated component.

Consider the construction of the structure function for the simple service system (Fig.3) using the FDT (Fig.5) that has been induced based on the monitored data for this system (Table 3). All possible component states (all state vectors) must be used for the calculation of the system performance level using FDT to form the decision table (structure function). The vector states are represented in the top three rows in Table 8.

Note that according to the FDT in Fig.5, the system performance level for all possible vector states can be defined as 0 with the confidence of 0.433, 1 with the confidence of 0.475 and 2 with the confidence of 0.092. However, these confidences are under the threshold value of $\beta=0.90$. The analysis of the state vectors is therefore implemented based on the FDT.

For example, assume that the state vector is $x = (0\ 0\ 0)$. Analysis based on the FDT starts with the attribute A_3 (Fig. 5) that is associated with the third component. The value of this component state is 0 ($x_3 = 0$) for the specified state vector. Therefore, the branch for the attribute value $A_{3,0}$ is considered. According to this attribute value, the output attribute value (system performance level) is defined as 0 with the confidence of 0.930 without analysis of other attributes. Considering this, the system performance level for the state vector is $x = (0\ 0\ 0)$ is 0 (Table 8).

Let us consider the following state vector from Table 8: $x = (0\ 0\ 1)$. The value of the attribute A_3 is $A_{3,0}$ for this state vector. According to the FDT, the branch with the attribute value $A_{3,0}$ is analyzed. The output attribute can have: value 0 with the confidence of 0.240, value 1 with the confidence of 0.760 and value 2 with the confidence of 0.00. All these confidences are less than the threshold value of $\beta=0.90$. Therefore, a decision about the value of the system performance level is not possible and the analysis continues. The next analyzed attribute is A_1 . The estimation of this attribute is implemented using a branch with attribute value $A_{1,0}$ because the specified state vector includes $x_1 = 0$. Decision about the value of the system performance level cannot be made because confidence for all possible values of the system performance level are not higher than the threshold of $\beta=0.90$. The analysis is therefore continued with the next node for the attribute A_2 . According to the specified state vector, this attribute has the value $A_{2,0}$ because $x_2 = 0$. The branch of this value has the least node. Therefore, the value of the output attribute is defined as the value with the maximal confidence - that is 0.863 for value 0. Considering this, the system performance level for the specified state vector is $\phi(x) = 0$ (Table 8).

Analysis of other state vectors is similar and facilitates the acquisition of all possible values of the system performance

level in the form of the structure function that is defined in Table 8.

We shall discuss the situation for close attribute values. Now, one value for the output attribute with maximum confidence is chosen. This value is interpreted as the structure function value according to the definition (1). The construction of the structure function (1) is one of the main objectives of the FDT application in this paper. We are going to develop a proposed method for the structure function construction taking into account different degree of confidence for its state vectors. Degree of confidence of all possible values (or max confidence degree) of structure function will be saved in the form of matrix (or vector).

It is important to note that this method of constructing the structure function is based on FDTs facility to compute (restore) missing monitored data. It illustrates the comparison of the structure function in Table 4 and Table 8. The structure function performance level is not defined for four state vectors in Table 4 that has been obtained from the monitored data without any specific methods. In Table 8, the performance levels are computed for all state vectors based on the methods with the use of FDTs.

The representation of the system using the structure function allows the calculation of different indices and measures to estimate system reliability. For example, the probabilities of every system performance level can be computed (as shown in Table 9) for the simple service system (Fig.2) based on the structure function from Table 8.

Therefore, probabilities of the system performance can be calculated according to typical methods used in reliability engineering based on the structure function. Other measures can be computed by the structure function too. For example, importance measures for this system are defined according to the algorithms considered in [6, 7].

TABLE IX
PROBABILITIES OF SYSTEM PERFORMANCE LEVEL

System performance level	Probabilities of the system performance levels
0	$p_{3,0} + p_{1,0} \cdot p_{2,0} \cdot (p_{3,1} + p_{3,2} + p_{3,3})$
1	$(p_{1,0} \cdot p_{2,1} + p_{1,1} \cdot p_{2,0}) \cdot (p_{3,1} + p_{3,2}) + p_{1,0} \cdot p_{2,1} \cdot p_{3,3}$
2	$p_{3,3} \cdot (p_{1,0} \cdot p_{2,1} + p_{1,1} \cdot p_{2,0})$

B. Efficiency and accuracy investigation of the proposed method

In this section, we present a simple case study carried out in order to verify the modelling approach described in previous sections. We consider three systems: outline of an offshore electrical power generation system [1]; army battle plan [39]; a computer system with a memory subsystem subject to competing failure isolation and propagation effect [40]. The

TABLE VIII
SYSTEM STRUCTURE FUNCTION

Component states	x_1	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
	x_2	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
	x_3	0	1	2	3	0	1	2	3	0	1	2	3	0	1	2	3
System performance level	$\phi(x)$	0	0	0	0	0	1	1	1	0	1	1	1	0	1	2	2

structure function of each system can be defined based on the description of the system behavior in appropriate paper. All of these systems are MSS. Basic characteristic of these system are shown in Table 10.

TABLE X
INVESTIGATED SYSTEM CHARACTERISTICS

	Outline of an offshore electrical power generation system [1];	Army battle plan [39];	Computer system with a memory subsystem [40]
Numbers of the system	3	5	4
performance level			
Numbers of components	5	4	5
Numbers of state vectors (dimension)	243	108	512

We use structure functions of these systems to examine the efficiency and accuracy of our proposed method for the construction of the structure function based on uncertain data. Therefore, these structure functions must be transformed to ambiguous and incomplete specified data. In the proposed methods, two types of uncertainly are included. The first is ambiguous data values. Therefore, all integer values of number for components states and performance level have been transformed to values with possibilities. Because the values of components states and performance level are defined in these structure function, we indicate known value with the possibility 1.00 and other values with possibilities 0.00. For example, the component has 4 states. Its value is indicated as “2” in one of state vectors. In the table of repository, this value is added as “0” with possibility 0.00, “1” with possibility 0.00, “2” with possibility 1.00 and “3” with possibility 0.00. We can use the algorithm from [41] for transform data from numeric to fuzzy cases in other situations.

The second type of considered uncertainty in our proposed method is incomplete specified initial data. This incompleteness is modeled by the random deletion of some state vectors and assigned performance level value. The range of deleted states is changed from 5% to 90%.

Each transformed structure function can be interpreted as uncertain monitored data. We use this data to construct the structure function based on the use of FDT induction. The FDT is induced by the method presented in [34, 35, 38] and considered in section IV briefly. The structure function construction is implemented according to the concepts introduced in section V. As a result, a single or a small group of state vectors may be misclassified. Therefore, we have to estimate this misclassification using an error rate. The constructed structure function and initial complete and precise specified function are compared and the error rate is calculated as a ratio of erroneous values of the structure function to the dimension of unspecified part of the function.

The experiments have been iterated 1000 times for every system and fixed values of parameters α , β and numbers of unspecified state vectors. The unspecified state vectors are selected randomly in proportion to the dimension of the structure function from 5% to 90-95%. The best result has

minimal error rate. The results for investigated systems are shown in Table 11. The error rate is dependant on unspecified part of the initial data. This error increases significantly if the unspecified part is most than 80% for all investigated systems. And we can see insignificant growth of the error rate if the unspecified part is less than 10%. The values of parameters α and β have been defined for the best decision based on these experiments (Table 11).

TABLE XI

THE ERROR RATE FOR THE CONSTRUCTION OF THE STRUCTURE FUNCTION FOR THE OUTLINE OF OFFSHORE ELECTRICAL POWER GENERATION SYSTEM IN [1]; ARMY BATTLE PLAN IN [39]; COMPUTER SYSTEM WITH A MEMORY SUBSYSTEM IN [40]

Unspecified state vectors, in %	The system in [1] $\alpha=0.15/\beta=0.85$	The system in [39] $\alpha=0.5/\beta=0.5$	The system in [40] $\alpha=0.01/\beta=0.95$
5	0,0661	0,1940	0,2721
10	0,0637	0,1938	0,2487
15	0,0682	0,1926	0,2324
20	0,0661	0,1962	0,2255
25	0,0670	0,1930	0,2151
30	0,0662	0,1922	0,2109
35	0,0648	0,1941	0,2086
40	0,0663	0,1928	0,2065
45	0,0657	0,1949	0,1945
50	0,0659	0,1939	0,1976
55	0,0671	0,1949	0,1995
60	0,0673	0,1938	0,2029
65	0,0679	0,1942	0,2009
70	0,0700	0,1942	0,2096
75	0,0743	0,1936	0,2227
80	0,0813	0,1941	0,2301
85	0,0995	0,1945	0,2423
90	0,1465	0,1961	0,2629

The effect of parameters α and β on the error rate illustrates are depicted as graphs for: the offshore electrical power generation system [1] (in Fig. 6); army battle plan [39] (Fig. 7); computer system with a memory subsystem [40] (Fig. 8). These parameters are defined empirically. The best decision (with minimal error rate) is in blue color.

In addition, we provide the structure function reconstruction by the application of value that appears most frequently in the immediate neighborhood [42]. This method does not use specific analysis of the structure function and the error rate can be considered as maximal. The error rate of this method is denoted by red line and as “Maximal error” in graphs shown in Fig.6–8.

Therefore, analysis of the error rate for the proposed method for the construction of the structure function based on FDT shows that this method has good efficiency. This method is acceptable for the incomplete data and the incompleteness of initial data can be indicated from 10% to 85%. The constructed structure function by the proposed method has less

error rate than maximal error rate in interval of the incompleteness.

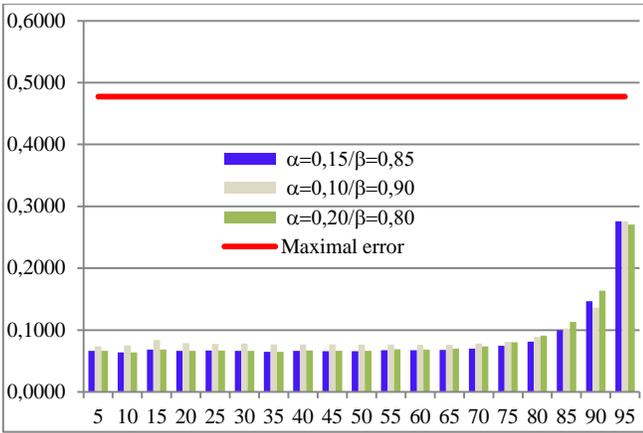


Fig.6. The error rate for the construction of the structure function for the outline of offshore electrical power generation system in [1] depending of different values of parameters α and β

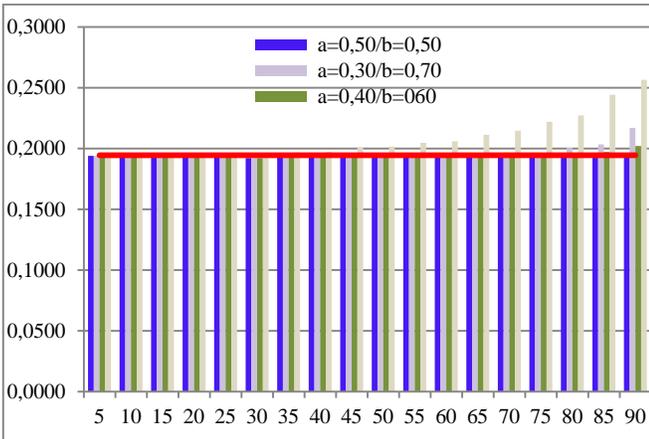


Fig.7. The error rate for the construction of the structure function for the army battle plan in [39] depending of different values of parameters α and β

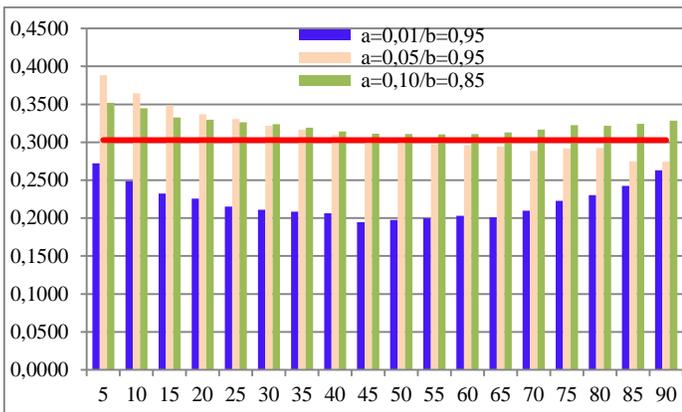


Fig.8. The error rate for the construction of the structure function for computer system with a memory subsystem in [40] depending of different values of parameters α and β

C. Future work

This paper has proposed a new method to construct a structure function based on uncertain data by application of

methodology of FDT and decision table. The conducted experiments show efficiency of this method. The accuracy of the method for examined system is acceptable. Future work will be devoted to investigation of more complicated cases. Systems with real ambiguous data values will be experimented on. In this case, all possibilities of values of components states and performance levels will not be equal to 1. The effect of this ambiguous data on error rate will be estimated.

The correlation of the system dimension and the method efficiency will be considered in the future too. Now, we can use the FDT application to forecast the dimension and specificity of the structure function. The construction of the structure function of 50 state vectors is possible based on 20-30 state vectors (40-60% of defined state vectors). The structure function construction based on 5-10 state vectors (10-20%) is possible too. But level of accuracy depends on the quality of this set of state vectors. It will be essential to continue the verification and validation of our proposed method with data sets of different properties and sizes

VI. CONCLUSION AND DISCUSSION

The new method for constructing the structure function has been proposed in this paper. This method allows obtaining a structure function based on uncertain data (for example, monitoring data). The term “uncertain” assumes uncertainties of two types. The first one is caused by the ambiguity of initial data. In this case, the system performance level and component states can be defined with any possibility. According to the typical definition of the structure function (1), performance level can have only one value for every state vector from the set $\{0, \dots, M-1\}$. However, the border between two neighboring values can be diffused in real applications. Both values can therefore, be indicated with any possibility. The proposed method takes such ambiguity into account and permits the performance level to be indicated using some values ranging from 0 to $M-1$ with a possibility that is considered in the subsequent steps of the method. The component states are indicated in a similar manner and the state of the i -th component is considered as a value ranging from 0 to m_i-1 with possibilities. For example, the monitored data in Table 2 are presented with the consideration of such ambiguity: every value is indicated with any possibility.

The second type of uncertainty deals with some state vectors missing from the initial data. In a practical application, it can be caused by the impossibility to obtain or indicate all possible combinations of system component states. For example, Table 3 does not include four values of the performance level because such combinations of component states have not been monitored. However, the proposed method has restored such values and permits the construction of a full structure function for all possible component states.

Ambiguity is considered and taken into account in the interpretation of the initial data as fuzzy data. This interpretation requires the use of mathematical methods of fuzzy logic for the analysis. In this paper, an FDT is used for system behavior modeling and construction of the system structure function. This mathematical method transforms

ambiguous and incomplete initial data to a correct decision [16, 33]. The induction of FDT is implemented based on cumulative information estimates [18] that takes into account the mathematical concept of entropy. These estimates are then adopted for the analysis of uncertain data. Therefore, the system structure function can be constructed using an FDT based on uncertain data and the FDT transforms incomplete specified data on system reliability/availability into a complete specified mathematical model that is the system structure function.

The main contribution of this paper is that we have developed a new and original method for the construction of a structure function based on incomplete specified and ambiguous data (monitoring data). This method facilitates the use of well know structure-function-based methods for MSS reliability estimation. The development of new specific methods for analysis of system reliability/availability based on uncertain monitoring data can be an alternative way.

REFERENCES

- [1] B.Natvig, *Multistate Systems Reliability Theory with Applications*, Wiley, New York, 2011.
- [2] E.Zio, "Reliability engineering: Old problems and new challenges", *Reliability Engineering and System Safety*, vol. 94, pp.125–141, 2009.
- [3] T.Aven, and B.Heide, "On performance measures for multistate monotone system", *Reliability Engineering and System Safety*, vol. 41, pp.259–266, 1993.
- [4] A.Lisnianski, and G.Levitin, *Multi-state System Reliability. Assessment, Optimization and Applications*. Singapore, SG: World Scientific, 2003.
- [5] R.E.Barlow, and B.Prochan, *Statistical Theory of Reliability and Life Testing, Probability Models*. Holt, Rinehart, and Winston, New York., 1975.
- [6] E.Zaitseva, "Importance analysis of a multi-state system based on multiple-valued logic methods," in *Recent Advances in System Reliability: Signatures, Multi-state Systems and Statistical Inference*, London: Springer, pp. 113–134, 2012.
- [7] M.Kvassay, E.Zaitseva, and V.Levashenko, "Minimal cut sets and direct partial logic derivatives in reliability analysis," in *Proc. ESREL*, Wroclaw, Poland, pp. 241–248, 2014.
- [8] V.Cutello, J.Montero, and J.Yanez, "Structure functions with fuzzy states", *Fuzzy Sets and Systems*, vol. 83, pp.189–202, 1996.
- [9] K.Kolowrocki, *Reliability of Large Systems*, Elsevier, 2004, 358 p.
- [10] R.E.Barlow, *Engineering reliability*. ASA and SIAM, Philadelphia, PA., 1998.
- [11] M.L.Gamiz, and M.D.Martinez Miranda, "Regression analysis of the structure function for reliability evaluation of continuous-state system," *Reliability Engineering & System Safety*, vol. 95, no. 2, pp.134–142, 2010.
- [12] M.L.Kansal, and S.S.Agarwal, "Fuzzy Based transformer Failure Analysis under uncertainty", in *Proc. ICROIT*, India, pp.1-5, 2014.
- [13] H.-Z.Huang, "Structural reliability analysis using fuzzy sets theory", *Eksplotacja i Niezawodność – Maintenance and Reliability*, vol.14, no. 4, pp.284–294, 2012.
- [14] Y.Liu, H.-Z.Huang, and D.Ling, "Reliability prediction for evolutionary product in the conceptual design phase using neural network-based fuzzy synthetic assessment", *Int. Journal of Systems Science*, vol. 44, no. 3, pp.545–555, 2013.
- [15] T.Aven, E.Zio, P.Baraldi, and R.Flage, *Uncertainty in Risk Assessment: The Representation and Treatment of Uncertainties by Probabilistic and Non-Probabilistic Methods*, Wiley, 200 p., 2014.
- [16] S.Mitra, K.M.Konwar, and S.K.Pal, "Fuzzy Decision Tree, Linguistic Rules and Fuzzy Knowledge-Based Network: Generation and Evaluation". *Trans. on Syst., Man Cybernetics — Part C: Applications and Reviews*, vol. 32, pp.328–339, 2002.
- [17] M.G.Tsipouras, T.P.Exarchos, and D.I.Fotiadis, "A methodology for automated fuzzy model generation," *Fuzzy Sets and Systems*, vol. 159, no 23, pp.3201–3220, 2008.
- [18] V.Levashenko, and E.Zaitseva, "Usage of New Information Estimations for Induction of Fuzzy Decision Trees," in *Lecture Notes in Computer Science LCNS2412*, Springer-Verlag, pp. 493–499, 2002.
- [19] H. Tanaka, L.T. Fan, F.S. Lai, and K. Toguchi, "Fault-tree analysis by fuzzy probability," *IEEE Trans. on Reliability*, vol.32, pp.453–457, 1983.
- [20] E.Zaitseva, and V.Levashenko, "Multiple-valued logic mathematical approaches for multi-state system reliability analysis", *Journal of Applied Logic*, vol. 11, no. 3, pp. 350–362, 2013
- [21] W.Kuo, and X.Zhu, *Importance Measures in Reliability, Risk, and Optimization: Principles and Applications*. Chichester, UK: Wiley, 2012.
- [22] T.Aven, and B.Heide, "Reliability and validity of risk analysis," *Reliability Engineering and System Safety*, vol. 94, no. 11, pp.1862–1868, 2009.
- [23] Y.Wang, and L.Li, "Effects of Uncertainty in Both Component Reliability and Load Demand on Multistate System Reliability," *IEEE Trans. on System, Man and Cybernetic – Part A: Systems and Humans*, vol.42, no. 4, pp. 958–969, 2012.
- [24] S.Virtanen, P.-E.Hagmark, and J.-P.Penttinen. "Modeling and analysis of causes and consequences of failures", in *Proc. RAMS*, Newport Beach, CA, USA, pp.506–511, 2006.
- [25] M.Li, J.Liu, J.Li, and B.U.Kim, "Bayesian modeling of multi-state hierarchical systems with multi-level information aggregation," *Reliability Engineering & System Safety*, vol.124, no. 4, pp.158–164, 2014.
- [26] O.O'Sullivan, A.Aboulafia, G.Iohom, B.D.O'Donnell, and G.D.Shorten, "Proactive Error Analysis of Ultrasound-Guided Axillary Brachial Plexus Block Performance," *Regional Anesthesia and Pain Medicine*, vol. 36, no.5, pp.502–507, 2011.
- [27] N.A.Stanton, and C.Baber, "Error by design: methods for predicting device usability," *Design Studies*, vol. 23, no. 4, pp.363–384, 2012.
- [28] D.Ley, "Approximating process knowledge and process thinking: Acquiring workflow data by domain experts," in *Proc. 2011 IEEE International Conference on Systems, Man, and Cybernetics*, pp. 3274 – 3279, 2011.
- [29] D.Yu, W.S. Park, "Combination and evaluation of expert opinions characterized in terms of fuzzy probabilities", *Annals of Nuclear Energy*, vol. 27, no 8, pp. 713–726, 2000.
- [30] X Chen, H. Hong, and E. Tamer "Measurement Error Models with Auxiliary Data," *Review of Economic Studies*, vol. 72, no.2, pp. 343–366, 2005.
- [31] J.R.Quinlan, "Simplifying decision trees", *Int. J. Man-Machine Studies*, vol. 27, pp. 221–234, 1987.
- [32] Y. Yuan, and M.J. Shaw, "Induction of Fuzzy Decision Trees," *Fuzzy Sets and Systems*, vol. 69, pp.125–139, 1995.
- [33] C.Olaru, and L.Whenkel, "A Complete Fuzzy Decision Tree Technique", *Fuzzy Sets and Systems*, pp.221–254, 2003.
- [34] V.Levashenko, E.Zaitseva, and S.Puuronen, "Fuzzy Classified based on Fuzzy Decision Tree," in *Proc. EUROCON*, Warsaw, Poland, pp.823–827, 2007.
- [35] V.Levashenko, E.Zaitseva, and S.Kovalik, *Decision Making based on the fuzzy data. Background, methods and algorithms*, LAP-Lambert Academic Publisher, 2014, 369 p (in Russian)
- [36] R. Krishnapuram, and J. Keller, "A possibilistic approach to clustering," *IEEE Trans. on Fuzzy Systems*, vol.1, pp.98–110, 1993.
- [37] R.Kruse, Ch. Doring, and M.-J. Lesot, "Fundamentals of Fuzzy Clustering," in *Advances in Fuzzy Clustering and its Applications*, eds. J.Valente de Oliveira and W. Pedrycz, Wiley, 434 p., 2007.
- [38] V.Levashenko, and E.Zaitseva, "Fuzzy decision trees in medical decision making support system," in *Proc. IEEE FedCSIS*, Wroclaw, Poland, pp.213–219, 2012.
- [39] R.A.Boedigheimer, and K.C.Kapur, "Customer-Driven Reliability Models for Multistate Coherent System," *IEEE Trans on Reliability*, vol.43, no.1, pp.46–50, 1994.
- [40] L.Xing, and G.Levitin, "Reliability of Multi-State System Subject to Competing Failures," in *Proc. Annual Reliability and Maintainability Symposium (RAMS)*, pp.1-7, 2011.
- [41] H.M. Lee, C.-M. Chen et al. "An efficient fuzzy classifier with feature selection based on fuzzy entropy," *IEEE Trans on Sys, Man, and Cyber, Part B: Cybernetics*, vol.31, pp. 426 - 432, 2001.
- [42] U.M. Braga-Neto, and E.R Dougherty, "Exact correlation between actual and estimated errors in discrete classification," *Pattern Recognition Letters*, vol.31, pp.407–412, 2010.

Elena Zaitseva received the M.S. degrees in computer science from the Belarus State University of Informatics and Radioelectronics (BSUIR), in 1989 and the Ph.D. degree in computer science from BSUIR, in 1994. From 1996 to 2004, she was an Assistant and Associated Professor with the Belarus State Economic University, Minsk, Belarus. Since 2004, she has been an Assistant Professor with the Department of Informatics, University of Zilina, Slovakia. Now she is a Professor at this Department.

She is the author of three books, more than 60 articles in Scopus and 6 inventions. Her research interests include Reliability Analysis of Multi-State System and its application in Health Care. She is Chair of the Technical Committee of European Safety and Reliability Association. Elena Zaitseva is the member of many Program Committee of International Conferences and Editorial Boards of several International Journals.

Vitaly Levashenko received the M.S. degree in computer science from the Belarus State University of Informatics and Radioelectronics (BSUIR), in 1991 and the Ph.D. degree in computer science from BSUIR, in 1997. From 1998 to 2004, he was an Associated Professor with the Belarus State Economic University, Minsk, Belarus. Since 2004, he has been an Assistant Professor with the Department of Informatics, University of Zilina, Slovakia. He is a Professor at this Department from 2013.

He is the author of three books, more than 50 articles in Scopus. His research main scientific and educational interests include Multiple-Valued and Fuzzy Logics, Knowledge Discovery in Databases, and Decision Making Support Systems in medical applications.