Application of Structure Function in System Reliability Analysis based on Uncertain Data

Vitaly Levashenko, Elena Zaitseva, Miroslav Kvassay

University of Zilina, Department of Infromatics,
Univerzitna 8215/1, 010 26, Zilina, Slovakia

{elena.zaitseva, vitaly.levashenko,
miroslav.kvassay}@fri.uniza.sk

Abstract. An important step in reliability evaluation of any system is selection of an appropriate mathematical representation. One of the possible mathematical representations is structure function that expresses dependency of system state on states of its components. This function must be completely specified for reliability evaluation of the analyzed system. The structure function is constructed based on complete information about the system structure and possible components states. However, there are a lot of practical problems when the complete information is not available because data from which it can be derived cannot be collected. As a rule, other mathematical representations and methods for evaluation of system reliability are used in these situations. In this paper, we propose a new method for construction of the structure function from uncertain or incomplete data. This method is developed based on application of Fuzzy Decision Tree.

Keywords: Fuzzy Decision Tree, Multi-State System, Structure Function, Uncertainty

Key Terms. Reliability, Model, Approach, Methodology, ScientificField

1 Introduction

As has been shown in paper [1], selecting a mathematical representation of an analyzed system is an important step in reliability analysis. Depending on the number of performance levels, two types of models can be recognized. These models are named as *Binary-State Systems* (BSSs) and *Multi-State Systems* (MSSs).

A BSS admits only two states in investigation of the system and its components: perfect functioning and complete failure. However, in practice, many systems can go through different performance levels between these two extreme states [1, 2]. A MSS is a mathematical model that is used to describe such systems since it allows defining more than two levels of performance [2, 3, 4].

There are different types of mathematical representations of a system. In reliability engineering, structure function, fault trees, reliability block diagrams, Markov models

ICTERI 2016, Kyiv, Ukraine, June 21-24, 2016

and Petri nets are typically used for the mathematical representation of real systems under study. Historically, mathematical models based on the structure function have been proposed firstly. In this case, a system is modeled as a mapping that assigns system state to all possible combinations of component states. The system performance level is known based on the states of all its components. This interpretation of the structure function supposes the exact definition of all possible states of the system and its components. Therefore, any uncertainty cannot be considered and taken into account. However, this indicates that methods based on the structure function approach have some difficulties in application on real-world problems because, as a rule, data about behavior of such systems are uncertain. Two approaches can be used to solve this problem.

The first of them is development of a new model that takes uncertainties into account [5, 6, 7]. The application of a new mathematical model leads to a development of new mathematical methods for the analysis of this model. The second solution is to use one of the traditional models and develop new methods for construction of the structure function that will take uncertainties of the initial data into account.

Specifics of the uncertainty have to be analyzed before the development of the new method for the structure function construction based on the uncertain data. There are different factors of uncertain data. In our investigation, we will take into account two of them. The first are ambiguity and vagueness of initial data. It means that initial data about the system operation are collected based on (a) measurement that can be inaccurate and with an error or (b) experts that can have different opinions on one situation. Therefore, values of states of the components or system performance level cannot be indicated as exact (integers). Ambiguity and vagueness in a real system have been studied using the probability theory. However, it is worth pointing out that some uncertainties that are not random in nature can play important roles in construction of the structure function [5, 6, 8]. The fuzzy logic makes it possible to define the structure function in a more flexible form for such data than the probabilistic approach. So, non-exact values are the first factor of the uncertainty of initial data, and it can be expressed using fuzzy values.

Secondly, situations in which it is impossible to indicate some values of the system components states or performance level can exist. For example, it can be very expensive, or it needs unacceptable long time. This implies that some information about the system behavior can be absent. Therefore, the data are incomplete.

Based on the previous text, we have a task of construction of exact and completely specified structure function based on uncertain and incomplete data, what is a typical problem of Data Mining [9]. One of the approaches used for solving this problem is application of *Fuzzy Decision Trees* (FDTs), which are widely used in Data Mining for analysis of uncertain data and decision making in ambiguities [10, 11].

In this paper, we propose a method based on the application of an FDT for construction of the structure function. FDTs allow taking into account uncertainties of two types. The first of them is ambiguity of initial data. This can occur when it is expensive to obtain all data about real system behavior, or there are poorly documented data. This type of uncertainty is covered by fuzzy values in an FDT. It means that initial data can be defined and interpreted with some possibility and might not be exact. The second type of uncertainty agrees with incompletely specified initial data. As a rule, if the exact values of the actual data about the system behavior cannot be determined, we need to rely on more data to get additional information necessary to correct the used theoretical model [6, 12]. An FDT allows reconstructing these data with different levels of the confidence [10, 11].

This paper is structured as follows. Section 2 discusses the concept of the structure function. The principal steps of the proposed method are considered in sections 3 – 5. These steps are Collection of data into a repository (section 3), Representation of the system model in the form of an FDT (section 4), and Construction of the structure function based on the FDT (section 5).

2 Structure function of the system

The *structure function* as a mathematical model was introduced in reliability engineering as one of the firsts [13]. This function captures the relationships between components of the system and the system itself in such a way that the state of the system is known based on the states of its components through the structure function.

Let us suppose that the system can be divided into n components (subsystems). A state of each component can be denoted by a random variable x_i that can be in one of m_i possible values. This variable takes value 0 if the component fails and one of values $1,\ldots, m_i -1$ if the component works satisfactorily.

Let us denote the structure function as $\phi(\boldsymbol{x})$. Then it agrees with the next map:

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

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

Next, let us suppose that the system is coherent. This means: (a) the system structure function is monotone: $\phi(x_i, \boldsymbol{x}) \leq \phi(x_j, \boldsymbol{x})$ for any $x_i \leq x_j$; and (b) there are no irrelevant components in the system.

Every system component is characterized by the probabilities of individual states:

$$p_{i,s} = \Pr\{x_i = s\},\ s = 0,\ldots, m_i -1 . \tag{2}$$

Please note that the structure function of MSS (1) is transformed into the structure function of BSS if $m_i = M = 2$.

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 [3]:

$$A_j = \Pr\{\phi(\boldsymbol{x}) = j\},\ j = 0,\ldots, m_i -1 . \tag{3}$$

The structure function also allows calculating the boundary system states [14], minimal cut/path sets [15] and importance measures [16]. However, defining structure function as equation (1) for a real application can be a difficult problem.

As a rule, the structure function can be defined as a result of the system structure analysis or based on expert data [12, 17]. In system structure analysis, the system is interpreted as a set of components (subsystems) with correlations. These correlations can be defined by functional relations that are interpreted as the structure function (1). However, there are many structure-complex systems for which correlations and/or connections of components are hidden or uncertain (e.g. power systems, network systems). As a rule, other methods are used in reliability estimation for such systems [5, 18]. Construction of a structure function based on the expert data requires special analysis and transformation of initial data [12, 19]. We suggest the new method for construction of the structure function (1) that is based on the application of an FDT.

In terms of Data Mining, the structure function can be interpreted as a table of decisions [9, 20], where state vector $\boldsymbol{x} = (x_1,\ldots, x_n)$ is interpreted as a set of input attributes and value of the structure function as an output attribute. This table of decisions can be constructed based on an FDT for all combinations of the input attributes. So, values of the structure function can be defined for all combinations of component states using the FDT: component states are interpreted as FDT attributes, and the structure function value agrees with one of M values (classes) representing system performance levels. The FDT is inducted based on some samples (not all) of the inputs and output attributes. In case of construction of the structure function, the samples are state vectors with the corresponding function value. These samples have to be collected as initial information about the system.

The method proposed in this paper includes the following steps:

- collection of data into 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 a decision table that is created by inducted FDT.

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

3 Data repository construction

Collection of data in the form of a repository is provided by the monitoring of values of system component states and system performance level. This repository can be presented in the form of a table where the columns agree with the input and output attributes. The number of the input attributes is n and the i-th has m_i possible values (the i-th column includes m_i sub-columns). Every row contains a real sample of components states and the corresponding system performance level.

For example, let us consider the offshore electrical power generation system presented in [2]. The purpose of this system (Fig. 1) is to supply two nearby oilrigs with electric power. The system includes 3 generators: two main generators A_1 and A_3, and standby generator A_2. Both main generators are at oilrigs. In addition, oilrig 1 has generator A_2 that is switched into the network in case of outage of A_1 or A_3. The control unit U continuously supervises the supply from each of the generators with automatic control of the switches. If, for instance, the supply from A_3 to oilrig 2 is not sufficient, whereas the supply from A_1 to oilrig 1 is sufficient, U can activate A_2 to supply oilrig 2 with electric power through the standby subsea cables L. This implies that the system consists of 5 relevant components (n = 5): generators A_1, A_2, and A_3, control unit U, and the standby subsea cables L. Furthermore, according to the description of the system activity in [2], we assume that the system and all its components have 3 states/performance levels (M = 3 and m_i = 3, for i = 1,…,5). Next, let us denote variables defining states of the system components in the following way: main generators A_1 and A_3 as x_1 and x_3 respectively, standby generator A_2 as x_2, and control unit U and standby subsea cables L as x_4 and x_5 respectively.

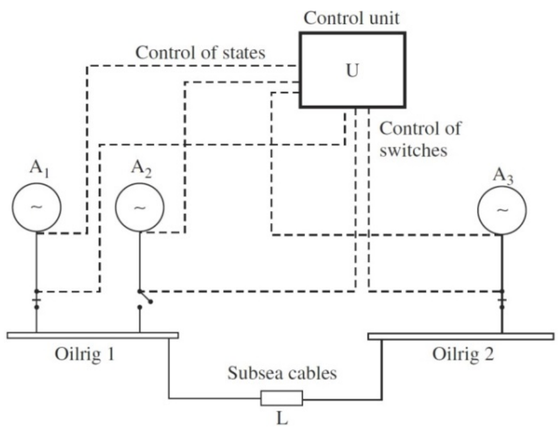

Fig. 1. Outline of the offshore electrical power generation system [2]

Let us suppose monitoring of the offshore power generation system that allowed collecting 108 (from 243 possible) samples of the system behavior. Some of them are shown in Table 1. The monitoring of this system permitted obtaining information about some combinations of component states and the corresponding performance levels of the system. However, this information is not complete because the data from the real monitoring are uncertain. This uncertainty is caused by the ambiguity of classification of component states and system performance levels into classes of exact values [12, 20]. Therefore, special type of data is used to define values of the input and output attributes in the repository. These data can be interpreted as quasi-fuzzy data that describe occurrence of every value of every attribute with some possibility ranging from 0 to 1. For example, the first row in Table 1 indicates the nonworking (x_1 = 0) and insufficient (x_1 = 1) states of generator A_1 with possibility of 0.8 and 0.2 respectively, while the possibility of the working state (x_1 = 2) is 0. In case of stable generator A_2, the state is indicated as nonworking (x_2 = 0) with possibility of 0.8 and as other values (x_2 = 1 and x_2 = 2) with possibilities of 0.1. State of main generator A_3

is nonworking ($x_3 = 0$) with possibility 0.7 and insufficient ($x_3 = 1$) or working ($x_3 = 2$) with possibilities 0.2 and 0.1 respectively. States of control unit U are defined as $x_4 = 0$ with possibility 0.8, $x_4 = 1$ with possibility 0.2 and $x_4 = 2$ with possibility 0. Only 2 of 3 states of the standby subsea cables L are relevant in this case because possibility of state $x_5 = 2$ is 0. The relevant states have possibilities 0.7 for $x_5 = 0$ and 0.3 for $x_5 = 1$. The system state is interpreted as a failure for this components states with the possibility 0.7 ($\phi(\boldsymbol{x}) = 0$) and as the sufficient state ($\phi(\boldsymbol{x}) = 1$) with the possibility 0.3, while the state of perfect operation ($\phi(\boldsymbol{x}) = 2$) is not indicated since its possibility is 0.

Table 1. Data obtained based on the monitoring of the offshore electrical power generation system

No	x_1			x_2			x_3			x_4			x_5			$\phi(\boldsymbol{x})$		
	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2
1	0.8	0.2	0.0	0.8	0.1	0.1	0.7	0.2	0.1	0.8	0.2	0.0	0.7	0.3	0.0	**0.7**	**0.3**	**0.0**
2	0.8	0.1	0.1	0.7	0.1	0.2	0.6	0.2	0.2	0.8	0.2	0.0	0.0	1.0	0.0	**0.8**	**0.1**	**0.1**
3	1.0	0.0	0.0	0.7	0.3	0.0	0.9	0.1	0.0	0.0	0.9	0.1	0.7	0.2	0.1	**1.0**	**0.0**	**0.0**
4	0.8	0.1	0.1	0.8	0.1	0.1	0.0	0.9	0.1	0.1	0.9	0.0	0.8	0.1	0.1	**0.2**	**0.7**	**0.1**
5	0.7	0.2	0.1	1.0	0.0	0.0	0.2	0.7	0.1	0.1	0.6	0.3	0.1	0.9	0.0	**0.0**	**0.7**	**0.3**
6	1.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0	0.7	0.2	0.1	0.5	0.3	0.2	**0.7**	**0.2**	**0.1**
7	0.8	0.1	0.1	0.2	0.6	0.2	0.8	0.1	0.1	0.0	0.0	1.0	0.8	0.1	0.1	**0.6**	**0.2**	**0.2**
8	1.0	0.0	0.0	0.0	0.9	0.1	0.0	0.9	0.1	0.0	0.9	0.1	0.7	0.3	0.0	**0.0**	**0.6**	**0.4**
9	0.8	0.1	0.1	0.0	0.9	0.1	0.1	0.8	0.1	0.0	0.9	0.1	0.2	0.7	0.1	**0.1**	**0.6**	**0.3**
10	0.7	0.3	0.0	0.1	0.8	0.1	0.0	0.1	0.9	0.8	0.1	0.1	0.8	0.2	0.0	**0.7**	**0.1**	**0.2**
11	0.7	0.2	0.1	0.0	0.9	0.1	0.0	0.1	0.9	0.0	0.1	0.9	0.7	0.2	0.1	**0.0**	**1.0**	**0.0**
12	0.1	0.6	0.3	0.8	0.2	0.0	0.2	0.8	0.0	0.1	0.7	0.3	0.0	1.0	0.0	**0.1**	**0.6**	**0.5**
13	0.2	0.8	0.0	0.7	0.3	0.0	0.1	0.1	0.8	0.1	0.6	0.3	1.0	0.0	0.0	**0.0**	**0.7**	**0.3**
14	0.1	0.8	0.1	0.2	0.7	0.1	0.0	1.0	0.0	0.3	0.6	0.4	0.9	0.1	0.0	**0.2**	**0.8**	**0.0**
15	0.0	0.2	0.8	0.9	0.1	0.0	0.2	0.8	0.0	0.0	0.1	0.9	0.0	0.1	0.9	**0.0**	**0.6**	**0.4**
16	0.0	0.1	0.9	1.0	0.0	0.0	0.0	0.1	0.9	0.1	0.6	0.3	0.0	0.2	0.8	**0.2**	**0.5**	**0.3**
17	0.0	0.2	0.8	0.1	0.6	0.3	0.2	0.5	0.3	0.2	0.7	0.1	0.0	0.3	0.7	**0.1**	**0.1**	**0.8**
18	0.0	0.2	0.8	0.2	0.7	0.1	0.0	0.2	0.8	1.0	0.0	0.0	1.0	0.0	0.0	**0.7**	**0.0**	**0.3**
19	0.0	0.1	0.9	0.0	0.2	0.8	0.2	0.7	0.1	0.0	0.1	0.9	0.2	0.7	0.1	**0.0**	**0.1**	**1.9**
	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	**…**	**…**	
108	0.0	0.0	1.0	0.0	0.1	0.9	0.0	0.1	0.9	0.0	0.1	0.9	0.0	0.8	0.2	**0.0**	**0.0**	**1.0**

The data obtained based on the monitoring and presented in Table 1 can be interpreted as fuzzy data [21]. The possibilities of individual states of the system components and of the system correspond to membership functions of fuzzy data.

The data obtained based on the monitoring of the offshore electrical power generation system are incompletely specified because we have 108 of all 243 combinations of components states. Traditional mathematical approach for system reliability analysis based on the structure function cannot be used in this case. Therefore, construction of structure function (1) based on incomplete data requires a special transformation and development of new methods. In this paper, we suggest the new method for construction of the structure function based on an FDT. This method allows reducing indeterminate values and obtaining a completely specified structure function.

4 Construction of FDT for representation of system

A decision tree is a formalism for expressing mappings of input attributes (components states) to output attribute/attributes (system performance level), consisting of an analysis of attribute nodes (input attributes) linked to two or more sub-trees and leafs or decision nodes labeled with classes of the output attribute (in our case, a class agrees with a system performance level) [21]. An FDT is one of the possible types of decision trees that permit operating with fuzzy data (attributes) and that use methods of fuzzy logic. Construction of a structure function assumes manipulation with real data, but the analysis of these data is implemented based on the methods of fuzzy logic the data are uncertain [22, 23]. The uncertainty may be present in obtaining numeric values of the attributes (system components states) or in obtaining the exact class (system performance level) where the instance belongs to.

There are different methods for inducting an FDT [10, 22, 24]. An FDT induction is implemented by the definition of the correlation between n input attributes $\{A_1,\ldots, A_n\}$ 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 defined by a state vector are input attributes. Each input attribute (component state) A_i $(1 \le i \le n)$ is measured by a group of discrete values ranging from 0 to m_i -1, which agree with the values of states of the i-th component: $\{A_{i,0},\ldots, A_{i,j},\ldots, A_{i,m_i\text{-}1}\}$. An FDT assumes that the input set $\mathbf{A} = \{A_1,..., A_n\}$ is classified as one of the values of output attribute B. Value B_w of output attribute B agrees with one of the system performance levels and is defined as M values ranging from 0 to M -1 (w = 0,…, M -1). The correlation between the terminologies and basic concepts of FDTs and reliability analysis are shown in Table 2.

Table 2. Correlation between the terminologies of FDTs and reliability analysis

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

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

For example, let us consider input attributes $\mathbf{A} = \{A_1, A_2, A_3, A_4, A_5\}$ and the output attribute B for the offshore electrical power generation system in Fig. 1. This system is represented by 5 input attributes. Each input attribute is defined as: $A_i = \{A_{i,0}, A_{i,1}, A_{i,2}\}$, for $i = 1,\ldots, 5$, and the output attribute is $B = \{B_0, B_1, B_2\}$. The values of the input attributes and the output attribute are defined in Table 3. These values are obtained based on the data from Table 1 and are used for the FDT construction as a training test. The principal difference of Table 1 and 3 is interpretation of initial date as attributes. The induction of the FDT based on this training test can be implemented using some of methods for FDT induction [10, 22, 24]. We propose to induct the FDT using the method based on the *cumulative information estimates* proposed in [25]. These estimations allow inducting FDTs with various properties. Criteria for building non-ordered, ordered or stable FDTs, as well as, development of this method have been considered in [26].

Table 3. A training set for the FDT induction

No	A_1			A_2			A_3			A_4			A_5			B		
	$A_{1,0}$	$A_{1,1}$	$A_{1,2}$	$A_{2,0}$	$A_{2,1}$	$A_{2,2}$	$A_{3,0}$	$A_{3,1}$	$A_{3,2}$	$A_{4,0}$	$A_{4,1}$	$A_{4,2}$	$A_{5,0}$	$A_{5,1}$	$A_{5,2}$	B_0	B_1	B_2
1	0.8	0.2	0.0	0.8	0.1	0.1	0.7	0.2	0.1	0.8	0.2	0.0	0.7	0.3	0.0	**0.7**	**0.3**	**0.0**
2	0.8	0.1	0.1	0.7	0.1	0.2	0.6	0.2	0.2	0.8	0.2	0.0	0.0	1.0	0.0	**0.8**	**0.1**	**0.1**
3	1.0	0.0	0.0	0.7	0.3	0.0	0.9	0.1	0.0	0.0	0.9	0.1	0.7	0.2	0.1	**1.0**	**0.0**	**0.0**
4	0.8	0.1	0.1	0.8	0.1	0.1	0.0	0.9	0.1	0.1	0.9	0.0	0.8	0.1	0.1	**0.2**	**0.7**	**0.1**
5	0.7	0.2	0.1	1.0	0.0	0.0	0.2	0.7	0.1	0.1	0.6	0.3	0.1	0.9	0.0	**0.0**	**0.7**	**0.3**
6	1.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0	0.7	0.2	0.1	0.5	0.3	0.2	**0.7**	**0.2**	**0.1**
7	0.8	0.1	0.1	0.2	0.6	0.2	0.8	0.1	0.1	0.0	0.0	1.0	0.8	0.1	0.1	**0.6**	**0.2**	**0.2**
8	1.0	0.0	0.0	0.0	0.9	0.1	0.0	0.9	0.1	0.0	0.9	0.1	0.7	0.3	0.0	**0.0**	**0.6**	**0.4**
9	0.8	0.1	0.1	0.0	0.9	0.1	0.1	0.8	0.1	0.0	0.9	0.1	0.2	0.7	0.1	**0.1**	**0.6**	**0.3**
10	0.7	0.3	0.0	0.1	0.8	0.1	0.0	0.1	0.9	0.8	0.1	0.1	0.8	0.2	0.0	**0.7**	**0.1**	**0.2**
11	0.7	0.2	0.1	0.0	0.9	0.1	0.0	0.1	0.9	0.0	0.1	0.9	0.7	0.2	0.1	**0.0**	**1.0**	**0.0**
12	0.1	0.6	0.3	0.8	0.2	0.0	0.2	0.8	0.0	0.1	0.7	0.3	0.0	1.0	0.0	**0.1**	**0.6**	**0.5**
13	0.2	0.8	0.0	0.7	0.3	0.0	0.1	0.1	0.8	0.1	0.6	0.3	1.0	0.0	0.0	**0.0**	**0.7**	**0.3**
14	0.1	0.8	0.1	0.2	0.7	0.1	0.0	1.0	0.0	0.3	0.6	0.4	0.9	0.1	0.0	**0.2**	**0.8**	**0.0**
15	0.0	0.2	0.8	0.9	0.1	0.0	0.2	0.8	0.0	0.0	0.1	0.9	0.0	0.1	0.9	**0.0**	**0.6**	**0.4**
16	0.0	0.1	0.9	1.0	0.0	0.0	0.0	0.1	0.9	0.1	0.6	0.3	0.0	0.2	0.8	**0.2**	**0.5**	**0.3**
17	0.0	0.2	0.8	0.1	0.6	0.3	0.2	0.5	0.3	0.2	0.7	0.1	0.0	0.3	0.7	**0.1**	**0.1**	**0.8**
18	0.0	0.2	0.8	0.2	0.7	0.1	0.0	0.2	0.8	1.0	0.0	0.0	1.0	0.0	0.0	**0.7**	**0.0**	**0.3**
19	0.0	0.1	0.9	0.0	0.2	0.8	0.2	0.7	0.1	0.0	0.1	0.9	0.2	0.7	0.1	**0.0**	**0.1**	**1.9**
	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	**…**	**…**	
108	0.0	0.0	1.0	0.0	0.1	0.9	0.0	0.1	0.9	0.0	0.1	0.9	0.0	0.8	0.2	**0.0**	**0.0**	**1.0**

The FDT resulted from the training set presented in Table 3 has been inducted by application of the cumulative information estimates using the method in [24]. This FDT is presented in Fig. 2. The nodes of this FDT agree with the input attributes. Every node has 3 branches according to the values of the corresponding input attribute from the training test (Table 3). Every branch correlates with some values of the output attribute. The set of output attribute values in a branch is named as a leaf if the analysis finish and one of the values of the output attribute can be chosen according to algorithms proposed in [25, 26].

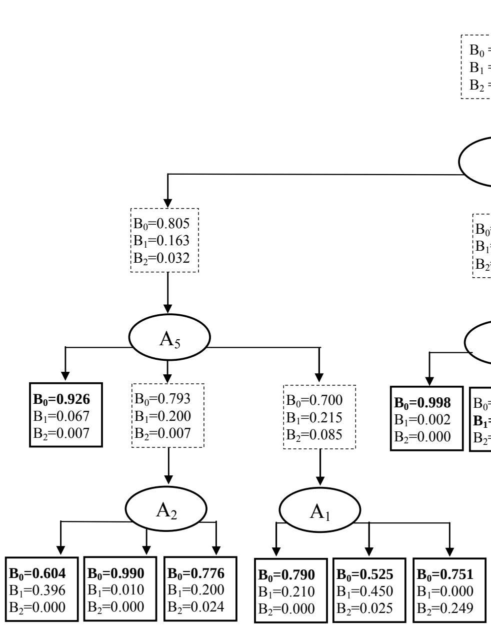

Fig. 2. Non-ordered FDT constructed based on the data obtained by the monitoring of the offshore electrical power generation system from Fig. 1

This FDT can be used for the analysis of all possible states of system components to construct the structure function of the offshore electrical power generation system. This process is considered below.

5 Construction of structure function based on FDT

According to [20], FDTs allow developing fuzzy decision rules or a decision table. A decision table contains all possible values of input attributes and the corresponding values of the output attribute that is calculated using the FDT. Such decision table agrees with the structure function. This implies that all possible combinations of values of the component states (all state vectors) have to be analyzed by the FDT to classify state vectors into *M* classes of the system performance levels.

Each non-leaf node is associated with an attribute $A_i \in \mathbf{A}$, or in terms of reliability analysis: each non-leaf node is associated with a component. The non-leaf node agreeing with attribute A_i has m_i outgoing branches. The *s*-th outgoing branch (s = 0,…, m_i -1) from the non-leaf node corresponding to attribute A_i agrees with state *s* of the i-th component ($x_i = s$). A path from the root to a leaf defines one or more state vectors (according to the values of the input attributes (component states) occurred in the path) for which the structure function takes value determined by the value of the output attribute. If any input attribute is absent in the path, all possible states have to be considered for the associated component.

Let us consider construction of the structure function of the offshore electrical power generation system from Fig. 1 using the FDT depicted in Fig. 2. All possible component states (all state vectors) have to be used for calculation of the system performance level by the FDT to form the decision table (structure function). Let us explain this idea for the first level of the FDT in more detail.

Preliminary analysis of the data obtained based on the monitoring (see Table 3) shows that possible values of the output attribute B are distributed as follows: value 0 – with confidence 0.493, value 1 – with confidence 0.209 and value 2 – with confidence 0.298. These values are implied by frequency of every output value in the training test (Table 3). Attribute A_3 is associated with the FDT root. So, analysis of the data starts from this attribute. This attribute has the following possible values: $A_{3,0}$, $A_{3,1}$, and $A_{3,2}$. Value $A_{3,0}$ of this attribute makes the output attribute B to be B_0 (the system is non-operational) with the confidence of 0.805. Other variants, B_1 and B_2, of output attribute B can be chosen with the confidence of 0.163 and 0.012 respectively. If the attribute A_3 has other values, i.e. $A_{3,1}$ or $A_{3,2}$, then the analysis is done similarly.

If the value of attribute A_3 is $A_{3,0}$, than the next attribute in the analysis is A_5, which have values $A_{5,0}$, $A_{5,1}$, and $A_{5,2}$. Value $A_{5,0}$ of this attribute agrees with a leaf representing the output attribute. Therefore, in this situation, the analysis is stopped and value of the output attribute is defined: value B_0 of attribute B should be chosen with the confidence of 0.926, and values B_1 and B_2 with confidences 0.067 and 0.007 respectively. Similarly, the process of the analysis of the non-ordered FDT continues for the other input attributes and their values.

Next, let us consider state vector $\boldsymbol{x} = (0,0,0,0,0)$. The analysis based on the FDT starts with the attribute A_3 (Fig. 2) that is associated with the 3-rd component. State of this component is 0 ($x_3 = 0$) for the specified state vector. Therefore, the branch for value $A_{3,0}$ of attribute A_3 value is taken into account. According to this value, the identification of the output attribute value (system performance level) has to continue through attribute A_5. According to the state vector, $x_5 = 0$, therefore, attribute A_5 has value $A_{5,0}$. Now, value of the output attribute can be indicated because the branch has a leaf. So, value of the output attribute is defined as 0 with the confidence of 0.926 without analysis of other attributes. Analysis of other state vectors is similar and allows obtaining all possible values of the system performance level in the form of the structure function. The analysis of all possible state vectors from $\boldsymbol{x} = (0,0,0,0,0)$ to $\boldsymbol{x} = (2,2,2,2,2)$ allows us to construct the complete structure function of the offshore electrical power generation system depicted in Fig. 1.

It is important to note that this method of construction of the structure function based on FDTs permits to compute (restore) data missing from the monitoring.

A representation of the system using the structure function allows calculating different indices and measures for estimation of system reliability. Probabilities of system performance levels (3) are one of them. Suppose that probabilities of the components states of the offshore electrical power generation system have values shown in Table 4. In this case, the probabilities of system performance levels are: A_2 = 0.73, A_1 = 0.20 and A_0 = 0.07. Other measures can be computed using the structure function too. For example, importance measures for this system can be calculated using the algorithms considered in [15, 27, 28].

Table 4. Components states probabilities

Component state, *s*	Probabilities				
	$p_{1,s}$	$p_{2,s}$	$p_{3,s}$	$p_{4,s}$	$p_{5,s}$
0	0.1	0.2	0.1	0.1	0.1
1	0.4	0.4	0.4	0.2	0.1
2	0.5	0.4	0.5	0.6	0.8

Let us present a simple case study to verify the modelling approach described in the previous sections. We use structure function of the offshore electrical power generation system to examine efficiency and accuracy of the proposed method for construction of the structure function based on uncertain data. Therefore, the structure function must be transformed to ambiguous and incompletely specified form. In the proposed methods, two types of uncertainty are included. The first is ambiguity of data values. Therefore, all integer values of components states and performance level have to be transformed to values with some possibilities. We can use algorithm from [29] to transform data from numeric to fuzzy cases in this case. The second type of the considered uncertainty in the proposed method is incompletely specification of initial data. This incompleteness is modeled by random deleting of some state vectors and the corresponding values of system performance levels. The range of deleted states was changed from 5% to 90%. Each transformed structure function can be interpreted as uncertain data obtained by the aforementioned monitoring. We used these data to construct the structure function based on the proposed methods using FDT induction. The FDTs were inducted based on the method presented in [22, 25]. The structure function construction was implemented according to the concept introduced in section 5. As a result, a single or a small group of state vectors might be misclassified. Therefore, we had to estimate this misclassification by the error rate. The constructed structure functions were compared with the exact specified function (it was defined at the beginning of the experiments), and the error rate was calculated as a ratio of wrong values of the structure function to the dimension of unspecified part of the function. The experiments were repeated 1000 times for every version of incompletely specified offshore electrical power generation system. The unspecified state vectors were selected randomly in proportion to dimension of the structure function from 5% to 85%. The results for the investigated system are shown in Table 5. The error rate depends on unspecified proportion of the initial data. This error increases essentially, if the unspecified part is most than 80%. This indicates that the proposed method is acceptable for large range of incompletely defined initial data and can be used for construction of the structure function based on incompletely specified data.

This method can be considered for special cases and types of initial data with application of other algorithms from Data Mining to improve the obtained results. For example, initial data can be obtained for similar samples and, in this case, special algorithms for pattern recognition and intelligent diagnosis can be used [30].

Table 5. The error rate for the construction of the structure function of the offshore electrical power generation system

Unspecified state vectors, in %	The error
5	0.0661
10	0.0637
15	0.0682
20	0.0661
25	0.0670
30	0.0662
35	0.0648
40	0.0663
45	0.0657
50	0.0659
55	0.0671
60	0.0673
65	0.0679
70	0.0700
75	0.0743
80	0.0813
85	0.0995
90	0.1465

6 Conclusion

The new method for constructing the structure function is proposed in this paper. This method allows obtaining a structure function based on incompletely specified data (for example, data obtained from some monitoring). The term “incompletely specified” assumes uncertainties of two types.

The first type of uncertainty deals with some state vectors missing from the initial data. In practical application, it can be caused by the impossibility to obtain or indicate all possible combinations of system component states.

These uncertainties are considered and taken into account in the interpretation of the initial data as quasi-fuzzy data. This interpretation requires 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 incomplete and uncertain initial data to a correct decision [10,24]. The induction of FDT is implemented based on cumulative information estimates [25] that take into account 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 incompletely specified data about system reliability/availability into a completely specified mathematical model that is known as the system structure function.

The second type of uncertainty results from ambiguity of initial data. In this case, the system performance level and components states can be defined with some possibilities. According to the typical definition of the structure function (1), performance level can have only one value for every state vector from set {0, …, M -1}. However, the boundary between two neighboring values can be diffused in real applications. Both such values can be therefore indicated with some possibility. The proposed method takes such ambiguity into account and permits indicating performance level using some values ranging from 0 to M -1 with a possibility that is considered in the next steps of the method and is not disregarded. 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 data in Table 1 are presented with consideration of such ambiguity: every value is indicated with some possibility.

7 Acknowledgment

This work is supported by the grant grants VEGA 1/0498/14 and VEGA 1/0038/16.

References

8 Zio E., Reliability engineering: Old problems and new challenges, Reliability Engineering and System Safety, 94, 125–141 (2009)
9 Natvig B., Multistate Systems Reliability Theory with Applications, Wiley, New York (2011)
10 Aven T., Heide B. (1993), On performance measures for multistate monotone system, Reliability Engineering and System Safety, Vol. 41, pp.259–266
11 Lisnianski A., Levitin G., Multi-state System Reliability. Assessment, Optimization and Applications. World Scientific, Singapore (2003)
12 Cutello V. Montero J., Yanez J., Structure functions with fuzzy states, Fuzzy Sets and Systems, 83, 189-202 (1996)
13 Huang H.-Z., Structural reliability analysis using fuzzy sets theory. Eksploatacja i Niezawodnosc – Maintenance and Reliability, 14 (4), 284–294 (2012).
14 Liu Y., Huang H.-Z., Ling D., Reliability prediction for evolutionary product in the conceptual design phase using neural network-based fuzzy synthetic assessment, International Journal of Systems Science, 44 (3), 545–555 (2013)
15 Aven T., Heide B., Reliability and validity of risk analysis, Reliability Engineering and System Safety, 94 (11), 1862–1868 (2009)
16 Witten I.H., Frank E., Hall M. A., Data mining : practical machine learning tools and techniques, Elsevier (2011)
17 Mitra S., Konwar K.M., and Pal S.K., Fuzzy Decision Tree, Linguistic Rules and Fuzzy Knowledge-Based Network: Generation and Evaluation. Journal of IEEE Trans. on Syst., Man Cybernetics — Part C: Applications and Reviews, 32, 328-339 (2002)
18 Tsipouras M. G.,. Exarchos T.P, Fotiadis D.I., A methodology for automated fuzzy model generation, Fuzzy Sets and Systems, 159(23), 3201-3220 (2008)

19 Aven T., Zio E., Baraldi P., Flage R., Uncertainty in Risk Assessment: The Representation and Treatment of Uncertainties by Probabilistic and Non-Probabilistic Methods, Wiley (2014)
20 Barlow R.E. Proschan B. Statistical Theory of Reliability and Life Testing, Probability Models. Holt, Rinehart, and Winston, New York (1975),
21 Zaitseva, E., Levashenko, V., Kostolny J., Kvassay, M., Direct partial logic derivatives in analysis of boundary states of multi-state system, 11th International Conference on ICT in Education, Research and Industrial Applications: Integration, Harmonization and Knowledge Transfer, pp. 535-549 (2015)
22 Kvassay M., Zaitseva E., Levashenko V. Minimal cut sets and direct partial logic derivatives in reliability analysis. In: Safety and Reliability: Methodology and Applications - Proceedings of the European Safety and Reliability Conference, pp. 241-248, CRC press (2015)
23 Kuo W., Zhu X., Importance Measures in Reliability, Risk, and Optimization: Principles and Applications. Chichester, Wiley (2012)
24 Wang Y. and Li L., Effects of Uncertainty in Both Component Reliability and Load Demand on Multistate System Reliability, IEEE Tranc. on System, Man and Cybernetic – Part A: Systems and Humans, 42(4), 958-969 (2012)
25 Li M., Liu J., Li J., Kim B.U., Bayesian modeling of multi-state hierarchical systems with multi-level information aggregation, Reliability Engineering & System Safety, 124(4), 158–164 (2014)
26 Stanton N.A., Baber C., Error by design: methods for predicting device usability, Design Studies, 23(4), 363-384 (2002)
27 Levashenko V., Zaitseva E., Kovalik S., Decision Marking based on the fuzzy data. Background, methods and algorithms, LAP-Lambert Academic Publisher (2014)
28 Quinlan J.R. (1987), Simplifying decision trees. International Journal of Manmachine Studies, 27, pp. 221-234.
29 Levashenko V., Zaitseva E., Puuronen S. (2007), Fuzzy Classified Based on Fuzzy Decision Tree., Proc. IEEE Int. Conf. on Computer as a tool (EUROCON 2007), Warsaw, Poland, pp.823 – 827.
30 Dong M., Kothari R. (2001), Look-Ahead Based Fuzzy Decision Tree Induction. IEEE Trans. on Fuzzy Systems, vol. 9, no.3, pp. 461-468.
31 Olaru C., Whenkel L. (2003), A Complete Fuzzy Decision Tree Technique. Fuzzy Sets and Systems, , pp.221-254
32 Levashenko V., Zaitseva E. (2002), Usage of New Information Estimations for Induction of Fuzzy Decision Trees. H. Yin et al. (Eds.): IDEAL 2002, Lecture Notes in Computer Science LCNS2412, Springer-Verlag, pp. 493-499.
33 Levashenko V., Zaitseva E. (2012), Fuzzy decision trees in medical decision making support system, Proc. of the IEEE Federated Conf. on Computer Science & Information Systems, (IEEE FedCSIS'12), September 9-12, Wrocław, Poland, pp.213-219.
34 Zaitseva E. (2012), Importance analysis of a multi-state system based on multiple-valued logic methods. In: Lisnianski A. and Frenkel I. (eds) Recent Advances in System Reliability: Signatures, Multi-state Systems and Statistical Inference, London: Springer, pp. 113–134
35 Artsiomchyk, Y. , Zhivitskaya, H. (2015), Improving the supply chain efficiency based on reliability analysis, Proc. Int. Conf. on Information and Digital Technologies, IDT 2015; Zilina; Slovakia,2015, pp. 16-21
36 Lee H.M., Chen C.-M. et all. An efficient fuzzy classifier with feature selection based on fuzzy entropy, IEEE Trans on Sys, Man, and Cyber, Part B: Cybernetics, vol.31, 2001, pp. 426 – 432
37 Subbotin S.A.(2014), The sample properties evaluation for pattern recognition and intelligent diagnosis, Proc. Int. Conf. on Digital Technologies, Zilina, Slovakia, pp. 321-332.