REVIEW OF THE HABILITATION THESIS OF MR. ING. JÁN RABČAN, Ph.D.

Title: **Development of common methods from reliability analysis and machine learning**

Reader: prof. Ing. Radim Briš, CSc., VSB – Technical University of Ostrava, Faculty of Electrical Engineering and Computer Science

The habilitation thesis of Mr. Ing. Ján Rabčan, Ph.D. deals with interconnection between reliability analyses and classification methods. Basic and widely used mathematical model in reliability analysis is the structure function. Classification is commonly used Machine Learning technique (ML). In the thesis a new approach for structure function construction is proposed which is based on classification. In other words, the structure function is interpreted as a classification process. The main goals of the work are demonstrated in pages 13-14: *i*) The elaboration of ML methods in reliability engineering for the analysis and evaluation of systems based on uncertain initial data, for the construction of a mathematical model of a system for its reliability analysis; and *ii*) The development of new methods for the "classifiers induction", in particular, attribute selection based on reliability evaluation of input attributes which allows the combination of the wrapper and embedded techniques.

The habilitation thesis is structured into three chapters. The first chapter deals with the reliability analysis of systems resulting from the structure function. The second chapter brings two new methods for feature selection. The third chapter provides two case studies where the developed methods are applied and demonstrated in selected real applications.

The first chapter contains 36 pages. In this chapter I appreciate the exhaustive and clearly arranged description of all possible forms of data uncertainty. Besides a possible presence of uncertainty the chapter brings the new method for structure function construction in the form of MDD (Multiple-Valued Decision Diagram) based on uncertain data. The Decision Tree (in fact Fuzzy Decision Tree-FDT, where fuzzy logic enables to manage an uncertain or imprecise data) is used for construction of the MSS structure function, which is further used for the calculation of diverse importance measures, what can be considered as one of novelties of this work. Originality of this work is in following: Fuzzy C-Means clustering is used for initial data fuzzification, CMI (Cumulative Mutual Information)-based method is used for FDT induction and MOM (Mean-of-maxima) method is used for defuzzification. This combination of methods showed to be very effective from the computational point of view.
This chapter also provides a case study that shows the use of the proposed method for the structure function construction.

Questions to chapter 1:

1. Page 30: "...the performance levels probabilities of the MSS 2-out-of-3 are computed as sum of all paths from rote node to specified terminal node". Is the probability P_1 well computed? In addition, Table 1.2 is not consistent with Table 1.3. Which system state is obtained by vector (1,1,0)?
2. Question to Case Study 1: In page 41 is mentioned that "there is the dataset of observations, which consists of 6 samples". The datasets are not introduced in the chapter. Can you describe in more detail, how look the datasets and how you

transformed the datasets to fuzzy data in Table 1.6? The same question is connected with illustrative example 2-out-of-3 system, page 39. What is the initial data used for fuzzification? Is it Table 1.3?

The second chapter contains 23 pages: At the beginning of this chapter feature selection methods are in detail recapitulated, advantages as well as disadvantages are thoroughly discussed. Novelty of this chapter is in the fact that two new methods for feature selection enabling the classifier induction are proposed. One of them (Chapter 2.2) is based on the use of the structure function in the form of MDD where features are step by step eliminated on the basis of importance analysis (IA) of input components (attributes). The significance of input components determined through IA is employed as a ranking procedure in the feature selection process.
The second method (Chapter 2.3) proposes a new method for feature selection based on attribute importance estimated from Fuzzy Decision Tree (FDT). The main novelty of this method is the use of FDT models which allows the handling of possible uncertainty occurring in data. The proposed method is based on the Mean Decrease in Impurity (MDI) measure. MDI here provides a rank of attributes which can be interpreted as the importance of system components, which is a specific type of importance measure. In the reliability context author differentiates between input attributes (as component states of a system) and output attribute (as the system performance).
Last part of the chapter 2 is devoted to 4 experiments. All of them confirmed that both proposed methods showed to be more effective in all of pursued aspects (classification accuracy, running time).

Questions to chapter 2:

3. Page 61: You say that defuzzification is a crucial process, for which you used "maximum membership principle". I miss a clarification why the method is optimal? Did you realized any experiment where you compared it with weighted average method, centroid method, or other?
4. Page 64: You mention that your method is based on Mean Decrease in Impurity (MDI) measure. I miss a close description or mathematical expression of this measure which seems to be a kind of classification error in first view. You mention that this measure is calculated in reference [113], but this internet-reference (https://...) is unavailable.

The third chapter has 20 pages. It consist of two case studies. The Case Study 1 (problem of timing of tracheostomy for COVID-19 patients) clearly confirmed conclusions of the research work that although some attributes with small values of *SI* (Structural Importance computed on the basis of FDT) are omitted, classification result remains practically unchanged. Needless to say that omitting of some attributes improves computational effectivity. In addition the proposed method for elimination of attributes can be used for different classifiers as well. The Case Study 2 shows similar conclusions that brings the use of the proposed method in context of attribute selection based on *SI*.

Other critical remarks:

- Page 13 and 28... not clear what means [Geh2018] and [Bry1986].
- Page 26... what is m (without index)? It should be M!
- Page 35...In the middle of the page is inconsistent sentence: "In this work, ..."
- Page 101...excessive number of authors!

- Page 102 … explain the reference [78]!

Conclusion

Mr. Ing. Ján Rabčan, Ph.D. has published his results on various journals, conferences and workshops. They were accepted by scientific community in the area of applied informatics. He is author or co-author of 50 publications on Scopus (including articles in highly recognized journals), having *h*-index 8 (excluding self-citations).

In the habilitation thesis Mr. Ing. Ján Rabčan, Ph.D. clearly demonstrated his ability to produce original creative research work. He proposed the novel approach for constructing the structure function of MSSs using incompletely specified and uncertain data. Two novel and effective methods for the determination of the most important attributes in the process of classification are proposed, first is based on IA and second on Tree Attribute Ranking.

The habilitation thesis fulfilled all important criteria and thus I recommend Mr. Ing. Ján Rabčan, Ph.D., after a successful defense of his work to receive the scientific-pedagogic degree "Docent" from the Faculty of management science and informatics, University of Zilina in Zilina.

prof. Ing. Radim Briš, CSc., Ostrava, January 5, 2024
VSB – Technical University of Ostrava, Electrical Engineering and Computer Science,
Department of Applied Mathematics,
17. listopadu 15, 708 33 Ostrava-Poruba