doc. Ing. Jaroslav MAJERNÍK, PhD.
Department of Medical Informatics and Simulator Medicine
Faculty of Medicine, Pavol Jozef Šafárik University in Košice
Trieda SNP 1, 040 11 Košice

REVIEW

of the habilitation thesis

of Ing. Ján Rabčan, PhD. on the topic

„Development of common methods from reliability analysis and machine learning"

The opponent's review on Mr. Ján Rabčan's habilitation thesis was drawn up based on the request of University of Žilina in Žilina, Faculty of management science and informatics from 5th December 2024. The assessment of Ján Rabčan's scientific and pedagogical contribution to the habilitation procedure is based only on the submitted habilitation thesis entitled "Development of common methods from reliability analysis and machine learning" and the information contained in the habilitation file, which were delivered to me together with the request for the preparation of the review.

The presented habilitation thesis is written in English in the range of 106 pages, including 129 cited sources of used literature, mostly from the last decade. The content of the habilitation thesis was divided by the author into three main chapters, which are preceded by an Introduction chapter. The focus and the objectives of the work are stated at the end of this Introduction. The first main chapter deals with elaboration of machine learning based methods with the aim to analyse and evaluate the systems based on uncertain initial data and using its mathematical model for its reliability analysis. The development of new methods for the classifiers induction, focused on attribute selection, are described in the second chapter of the presented thesis. The third main chapter describes two case studies that were used to analyse the capability of the proposed method to identify the most important attributes, to induct a new classifier and to evaluate its efficiency based on selected attributes. The manuscript of habilitation thesis is finalised with Conclusion and Discussion chapter.

Topicality of the work

Today, there is no doubt that the field of machine learning is applied to all areas of human activity in which it is possible to collect, analyse and provide some type of information. Using its products, we are able to process even bigger amounts of data, to reveal hidden information among thousands if records, to offer outputs in critical decision making procedures and many others. Here, the very challenging engineering oriented tasks include also the topics related

to the ways to ensure reliability of the methods or systems being evaluated as well as to the safety related issues. Thanks to the advances in reliability engineering, there were various methods developed, optimised and improved to be applicable on various datasets. Despite of very high levels of many recent data-driven approaches, the research trends in machine learning area include new and new projects dealing with ways how to increase the reliability and safety of the operated systems while using more effective data processing methods with shortened computational time and/or reduced demands on computing resources. The topic of the habilitation thesis, which describes the proposal of new methods and algorithms development intended for construction of mathematical model of a system and that is usable for reliability analysis as well as development of two new approaches to the selection of the most relevant attributes for uncertain data, fits into this trends and, from my point of view, is therefore highly relevant.

Contribution of the work and fulfilment of its goals

Mr. Ján Rabčan formulated the goals of his habilitation thesis into two main areas that are (i) elaboration of machine learning based 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) development of new higher reliable 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.

To realize the first goal, the construction of structure function of multi state systems using incompletely specified and uncertain data was proposed and realized. Based on the structure function the importance measures were calculated that served as background for identification of the most important components/attributes. To cover uncertainty issues, especially the implicit ones, the principles of fuzzy logic were incorporated. The advantage of the chosen approach was the construction of multi-valued decision diagram in the final step of the procedure. This novelty was verified on two case studies related to motorbike safety and fleet of drones. Both confirmed functionality and usability and proved efficiency in computational complexity and in accuracy of mathematical model too.

The second part of the research work was oriented on development of more effective and precise data preprocessing in machine learning tasks, particularly on development of feature or attribute selection methods for classifier induction. The developed methods combine advantages of filters and wrapper, so the final result lead to two new proposed methods that offers possibilities of dimension reduction and classification accuracy improvement. The first proposed method used reliability importance measures, while the second one was based on attribute importance estimation from fuzzy decision tree. The main forces to develop new methods lied on the challenges to reduce computational complexity and to maintain or even improve selection performance.

To evaluate the classifier importance, the mean decrease in impurity approach was proposed by the author. Here, the obtained classifiers are transformed into the structure functions. Thus the method is able to handle uncertain data too, including explicit as well as implicit uncertainties. From this point of view, it contributes to applicability in wide range of classification related issues. Two proposed methods were compared using four experiments performed on eight publicly available datasets (UCI Machine Learning Repository). These experiments showed the performance of models is generally better after the feature section, the feature selection has mostly positive effect on classification accuracy, the proposed methods outperform the wrapper approach in terms of speed – running time, and both proposed methods (based on SI and MDI) can improve classification accuracy.

The proposed methods were also investigated using two case studies. One study evaluated attributes for the timing of tracheostomy in COVID-19 patients, and the second one evaluated attributes for the level of cancer disease. The comparison of classifiers showed the fuzzy classifiers have slightly better metrics with best results of Fuzzy Decision Tree and thus it can be used in such types of clinical studies. The influence of changes in selected attributes on decision-making process also proved it is similar to the value of the structural importance measure.

A significant contribution of the habilitation work is also the presentation of the basic starting points for the design of its solutions, which, however, must be understood in the context of the current time, when current information and specific technological options and solutions are available. Such factors develop and change rapidly in the digital space. I positively assess that the author's proposal was created on the basis of individual existing methods and approaches as a critical evaluation and synthesis of currently available information and the best solutions.

Comments and questions

In terms of formality, the work is processed at a very good level. Nevertheless, I have a few comments from a formal point of view. I recommend to formulate the goals of the work in separate chapter or section and not as a part of the Introduction. In Nomenclature and Acronyms, I recommend to use an alphabetical order of individual items. Also, there are many acronyms missing in this list. Most likely, due to the final wording of the individual parts of the habilitation thesis, the author used or repeated several parts of the text, which is distracting if the same information is presented three or four times. Several resources cited in the text were not listed in the references. There is wrong numbering of the sections starting after section 1.4.4, and also in Case Study 2: Fleet of drones. The readability, references or descriptions of several figures are not sufficient or provided in the related text of the thesis or it is not complete. For example, the description of the figure 2.1., or readability of figures 3.4 and 3.5, or missing axes labels in figure 3.6 and 3.9. All the formal shortcomings and mistakes were marked directly in the text of the printed version of the thesis. However, these formalities and typos do not affect the scientific quality of the research work and the results achieved.

From the content point of view, I have the following questions regarding the submitted habilitation thesis:

- When reading the manuscript for the first time, I felt confused with the value 2 presented in x_3 component and its column with value 0 in the table 2.1. Then I realised the type of the system and the proposed method's principle, so it is most likely correct. Can you confirm or comment on this? Similarly, the computation of P_1 (presented on page 30) uses $p_{3,0}$ that leads to 0 (Figure 1.4) - is that correct?

- It is stated, that the use of multiple-value decision diagram, instead of truth table, allows the number of system components up to 100. Did you tried to use/test our methods with systems above 24 components (in some experimental trials)? And how it affected the structure function construction if yes?

- In Case Study 1, in which the attributes for the timing of tracheostomy in COVID-19 patients were evaluated, a set of 14 attributes out of 22 was selected for the final analysis. It is clear this reduction leads to reduction of the dimension of the structure function that correlates with the computational complexity, but can you explain how this selection was made or what were the selection criteria? Also, was the proposed method used with all the attributes and lead to the same results?

Conclusion

The submitted habilitation thesis provides an original and comprehensive overview of knowledge on the issues of machine learning and reliability analysis. After studying it, and habilitation file too, I can state that the author's professional focus contributes not only to the summarisation of knowledge base and to design, realization and verification of proposed novelties that can be used in practice on various systems, but also to the enrichment of pedagogical scope in the field of teaching applied informatics topics. Based on the evaluation of the work mentioned above, I conclude that the work meets the requirements for the habilitation procedure, and therefore I recommend that Ing. Ján Rabčan, PhD., after successful defence of habilitation thesis, and after approval of the scientific council of the faculty, be awarded the scientific-pedagogical title "Associate professor" – "docent" in Slovak.

Košice 9 January 2024

....
doc. Ing. Jaroslav Majerník, PhD.