Dissertation report by Prof. Ing. Tomáš Kozubek, Ph.D.
(repetition of the dissertation defence)

Ing. Issam Salman, Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

Thesis title: Bayesian Network for Medical Data Analysis

Objectives and their fulfilment, methods used

The multidisciplinary thesis combines data analysis, statistics, and artificial intelligence knowledge to develop new algorithms based on Bayesian networks and apply them to solve different synthetic problems, a problem based on a publicly available dataset and one real-world problem from biomedicine focused on heart attack mortality prediction.

The thesis has three main objectives:

1.   Learning the structure of Bayesian networks and Belief Noisy-Or models from incomplete datasets.
2.   Dealing with incomplete and imbalanced data for Chow-Liu, tree-augmented naive Bayesian (TAN), and selective TAN (STAN), which is a feature selection method for TAN.
3.   Applying machine learning methods for heart attack mortality prediction based on their features' values.

The content of the submitted work shows that all three objectives have been met and further improved in the updated version of the dissertation work, where chapter 3.5 was relevantly enhanced and extended. Publications in conference proceedings and journals support the text and respective results.

Changes in the updated thesis

The author added a List of symbols and application of Random Forest and SVM to the real data for comparison purposes. Some typos and mathematical notation were corrected. The last section of the main chapter was renamed to "Evaluation of a Novel Bayesian Network Model for the Classification of Heart Disease", enhanced and relevantly extended concerning recommendations of the thesis defence committee. The author used a public dataset and the original real data from biomedicine to evaluate the BN model; he analysed the impact of different nodes in the Bayesian network on the prediction results and identified factors that influence heart disease and mortality. The extended part has been submitted to a journal in Q2 concerning Scopus. The git link, the algorithm's complexity, and an explanation of the Wilcoxon test have also been added.

Language and formal comments

The thesis is divided into four chapters, including an introduction and a conclusion. The individual chapters are written clearly, and the text is supplemented with illustrations in the form of figures and tables.
The thesis is written in a readable manner. The original text contained many typos and mistakes in English, but the updated version is significantly better.

Publication activity of the student

The student has co-authored six publications listed in the thesis (5 on Scopus and five on WoS), of which three are in Q3-Q4 impacted journals and five are relevant to the thesis topic. The enhanced and extended part of the thesis has been submitted to the Applied Clinical Informatics journal, Q2, on Scopus.

Questions

Answers to all original questions were correctly incorporated into the updated text of the thesis.

Conclusion

This thesis describes several ingredients that had to be studied, implemented, tested and compared with alternative state-of-the-art methods.

The content of the submitted PhD thesis demonstrates that its objectives have been met. Therefore, I recommend that the doctoral candidate be admitted to the defence of his thesis and that, on successful completion, the degree of doctor be conferred.

Ostrava, 14.2.2024

prof. Ing.
Tomáš
Kozubek, Ph.D.