Supervisor's Report for the PhD Thesis Defense by Issam Salman on Bayesian Networks for Medical Data Analysis

---

The motivation of Issam Salman's thesis are applications of probabilistic graphical models in medicine. Particularly, in his thesis, he deals with the problem of learning Bayesian networks in the context of incomplete and imbalanced data, which is a common attribute of medical data. In comparison with other machine learning methods Bayesian networks possess good explanation abilities, which is a property very important in application domains such as medicine.

The thesis starts with a practical medical problem - predicting the mortality of patients with acute myocardial infarction. Various machine learning techniques are applied to this problem. The data used in the numerical experiments were collected in Czech and Syrian hospitals. A novel approach for learning from incomplete and imbalanced data is proposed. It is based on the idea of using subsets of data by selecting data sets where values are observed for a given subset of variables. These subsets of data may be different for different subsets of variables. This also implies that different subsets of data can be used to estimate a certain model parameter. This approach allows to use more data sets than it would be possible if all incomplete data sets were discarded, which allows to estimate model parameters more reliably. This part of the thesis is mainly experimental and concludes with experimental comparisons of different classifiers.

The main contribution of the thesis is the proposal of a new method for learning the Bayesian network structure from incomplete data. The proposed method can be considered as a data imputation method, i.e. a method that imputes missing data using a statistical model. In the thesis, the Gaussian mixture model is used for this purpose. After this step, a standard Bayesian network structure learning method is applied. Namely, it is a method that maximizes the Bayesian Information Criterion (BIC). The proposed method is compared with other approaches on different synthetic data generated from models corresponding to practical applications, e.g. so-called BNO models consisting of conditional probability tables of a special type called noisy-or. Finally, the proposed method is also applied to a medical data set. The experiments performed show that in most cases the proposed method is significantly better or comparable to other tested methods.

During his doctoral studies, Issam Salman has demonstrated a strong commitment to applied research and to completing his doctoral studies. He has shown good programming skills. He has collaborated with medical doctors abroad (in Syria and Germany). He was able to obtain data from hospitals and also received feedback from medical doctors on the resulting models. He was always ready to take advice and apply it in his work. He managed to publish his results in conference proceedings and scientific journals listed in Web of Science. His work has been cited in research papers by other authors. Five of his research

papers are registered in the SCOPUS publication database. To date, these papers have received 13 citations. Recently, Issam Salman prepared another journal paper based on the results obtained in the final stage of his Ph.D. study. This paper is currently under peer review.

Finally, I would like to say that I think Issam Salman's thesis is a valuable contribution to the field of Bayesian network learning. His research has also opened up new research questions that can be studied by him or by other researchers.

I recommend the dissertation for defense.

In Praha, December 11, 2023

Ing. Jiří Vomlel, Ph.D.
Ústav teorie informace a automatizace
Akademie věd České republiky