INSTITUTE OF COMPUTER SCIENCE

Academy of Sciences of the Czech Republic

Martin Holeňa

Pod Vodárenskou věží 2, 182 07 Praha 8, phone: +420 266052921, fax: +420 286585789, e-mail: martin@cs.cas.cz
web: www.cs.cas.cz/~martin

Review of the PhD thesis

Adaptive Testing using Bayesian Networks

by Martin Plajner

The reviewed thesis investigates adaptive testing based on Bayesian networks, focusing on the frequently occurring condition of monotonicity. The theoretical results are validated on three real-world datasets from testing student skills, including a large dataset from the Czech National Final High School Exam, containing answers of more than 20000 students.

The thesis is a collection icluding six published papers of which Martin Plajner is the first author, preceded by a comparatively short summarizing introduction. Five of the papers have been published in proceedings of conferences, workshops or seminars taking place in the years 2015–2019. The last paper has been published in the International Journal of General Systems, where it was submitted in December 2018 and accepted 10 months later.

In the first paper, written for the Bayesian Modeling Applications Workshop at the 31st Conference on Uncertainty in Artificial Intelligence in 2015, the authors show how Bayesian networks can be used for adaptive testing of human skills. They design several different Bayesian networks and compare them on data obtained from a paper test of mathematical knowledge of grammar school students.

The second paper, written for the Eight International Conference on Probabilistic Graphical Models in 2016, introduces a generic model for computerized adaptive testing. It then shows that specializations of that generic model include the item response theory, traditionally used for computerized adaptive testing, but also Bayesian networks and neural networks. I appreciate that the paper explicitly pays attention to monotonicity, which is often encountered in real-world problems.

Continuing research of Martin Plajner into monotonicity in Bayesian networks has been presented in the third published paper in the collection, for the European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty in 2017. I appreciate in particular the proposed gradient-descent-based learning algorithm that respects monotonicity constraints.

The fourth paper, written for the 20th Czech-Japan Seminar on Data Analysis and Decision Making in 2018, deals with an aspect of applying Bayesian networks to adaptive testing of human skills, namely with the methodology of question selection. For combining with Bayesian networks, the question selection methods of maximization of the expected entropy reduction, of the maximization of the expected skills variance, and of the maximization of the

expected question variance are proposed, and it is experimentally verified that each of them is more suitable for adaptive testing than the commonly used sequential selection.

The fifth paper, for the Workshop on Uncertainty Processing in 2018, reports further progress in the research into monotonicity in Bayesian networks, continuing the research reported in the third paper. I appreciate the novel learning algorithm using increasing penalization to lead the gradient descent towards the admissible area of monotonicity of the parameters.

The publication that I appreciate most among all the six in the collection is the paper published 2019 in the International Journal of General Systems. It elaborates in substantially more depth than all the other included publications the proposed novel method of learning monotone parameters and the results of investigating the effect of monotonicity on parameter learning and on the resulting model. A comprehensive comparison with other available methods shows that that the novel method outperforms them on small datasets and yields at least comparable results on large datasets.

From the methodological point of view, I have several questions to the PhD student:

1. The employed information gain criterium is one of the key criteria in active learning. Have also other important active learning criteria been considered?

2. In the employed neural networks, which method was used to find the optimal network topology?

3. What justifies the statement "NNs are not suitable for unsupervised learning"? Consider self-organizing maps or autoencoders.

4. Have the p-values obtained in the tests applied to the experiments in the fourth paper and in the last publication of the collection been corrected for multiple hypotheses testing? If yes, which correction method has been used?

From the formal point of view, I have two problems with the thesis:

5. Different papers in the collection are sometimes quite similar, sharing even several parts of the text. This definitely decreases the information value of the collection.

6. Apart from the six published papers, the collection contains also three unpublished texts: two unpublished and unreviewed texts uploaded to arXiv, and a study that the PhD student submitted, 4 years ago, to his doctoral state exam. These texts are not admissible as parts of a collection of publications, and I have not taken them into consideration in my review. However, they could be incorporated into the thesis introduction preceding the collection. As to the state exam study, this has actually become a standard.

In spite of the above mentioned methodological and formal flaws, however, the submitted thesis clearly shows that Martin Plajner is able to perform creative and systematic research, herewith fulfilling the requirements for a PhD degree.

Prague, December 13, 2020

Martin Holeňa