UNIVERSITY of GUELPH

COLLEGE of ENGINEERING AND PHYSICAL SCIENCES

SCHOOL OF COMPUTER SCIENCE

MSc Defence

Friday February 7, 2020 at 3:00 PM in Reynolds, Room 1101

Learning Non-Impeding Noisy-AND Tree Model Based Bayesian Networks From Data

Qian Wang

Chair: Dr. Xiaodong Lin
Advisor: Dr. Yang Xiang
Advisory Committee: Dr. Stefan Kremer
Non-Advisory Committee: Dr. Luiza Antonie

Abstract

Bayesian Networks (BNs) are a widely utilized formalism for representing knowledge in intelligent agents on partially observable and stochastic application environments, including medical diagnosis, equipment troubleshooting, crime scene investigation, and many others. When typical conditional probability tables are used in BNs, the space complexity of BNs is exponential on the number m of parent variables per variable. The time complexity of inference by such BNs is also lower-bounded exponentially by m. The non-impeding noisy-AND Tree (NAT) model-based BNs can significantly improve both space and time complexity above, rendering both complexity measures linear on m, for a wide range of sparse BN structures. In this thesis, the challenges and the background on BN structure search (such as global search and heuristic search), structure learning with scoring metrics (such as MDL and BDe), as well as a couple of local models (such as the noisy-OR, noisy-AND, and noisy-MAX), will be reviewed. This research will study learning NAT model-based BNs from data by applying the Minimum Description Length (MDL) principle considering the hidden causes involved, with the heuristic search method, to learning the NAT-modeled BNs. To be specific, the MDL score will be decomposed into several components to allow the NAT-enabled search. This research will also describe the strategies, algorithms, and time complexities for each involved method. Experiments will be conducted to illustrate the effects of the density constraints on the learned BNs and the efficiency and accuracy of the NAT-modeled BNs. This work will advance the field of BN structure learning with local models, by focusing on inequality constraints through the NAT model. The future practitioners and theoreticians will also benefit from the analyses and discoveries in this thesis. They can make tractable inferences from historical data based on NAT causal independence, especially when data satisfy high treewidth and low-density structure.