Constraint-Based Approach to Learning Bayesian Networks

Student: Zhi Peen Su, Supervisor: James Cussens

University of Bristol, Department of Computer Science

Introduction

Bayesian networks have been widely used in the area of AI, medical diagnosis, decision making processes, and more. However, learning Bayesian networks from data is considered to be an NP-hard problem. Researchers have come up with two main approaches for finding the best Bayesian network given the data. A score-based approach and a constraint-based approach. This project will primarily cover the constraint-based approach to learning Bayesian networks from data.

[1] https://en.wikipedia.org/wiki/Bayesian_network

[2] https://scholarworks.umass.edu/cgi/viewcontent.cgi?article=1174&context=open_access_dissertations

[3] Judea Pearl. Probabilistic Reasoning in Intelligent Systems. https://dl.acm.org/doi/10.5555/534975

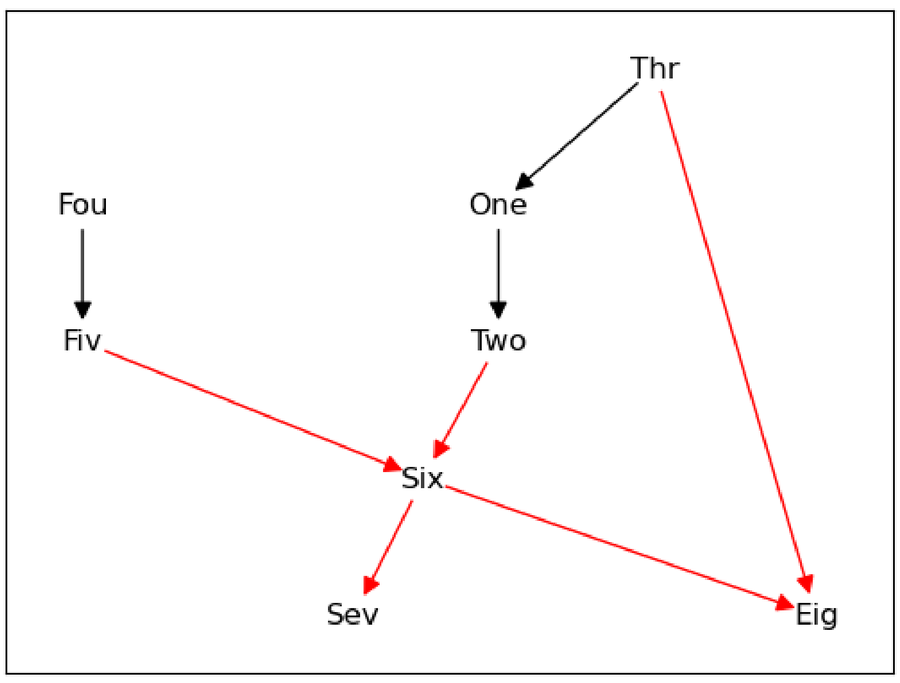
One	Two	Three	Four	Five	Six	Seven	Eight
2	2	2	2	2	2	2	2
1	1	0	0	0	1	0	1
1	0	0	0	0	0	1	0
0	0	1	0	0	0	1	0
1	0	1	0	0	0	1	0
1	0	1	0	0	0	0	1
1	0	1	0	0	0	1	1
1	0	1	0	0	0	1	1
1	0	1	0	0	0	0	1
1	0	1	0	0	0	0	1
1	0	1	0	0	0	1	0
1	0	1	0	0	0	1	1
1	0	0	0	0	0	1	0
0	0	0	0	0	0	1	0
1	0	1	0	0	0	1	1
0	0	0	0	0	0	1	0
0	0	0	0	0	0	1	0
0	0	0	0	0	0	1	0
0	0	0	0	0	0	1	0
1	0	0	0	0	0	1	0
0	0	1	0	0	0	1	1

asia_10000, small network (<10 nodes)

Score-Based Approach

Learning

Constraint-Based Approach



What is a Bayesian Network?

A Bayesian network [1] is a probabilistic graphical model which uses is represented by a directed acyclic graph (DAG). The nodes in the DAG represents variables, the arcs signifiy the existence of causal influences between the linked variables, and the strength of these influences are expressed by forward conditional probabilities [3].

What Are Constraints?

A constraint-based approach such as the PC algorithm starts by learning the skeleton (skeletal learning) of the graph by identifying constraints (independence assertions) of the form $(X \perp Y \mid Z)$, which means X is independent of Y given Z. An edge will be included in the skeleton if there is no identified independence assertion. [2]

Score-Based Approach

A score-based approach searches over all possible Bayesian network structures to find the best factorization of the joint distribution implied by the training data. The model selection criterion uses a penalized likelihood score such as the BDeu [3]. This has been implemented in pygobnilp.

Objective

As of now, utilizing only a score-based approach to learn Bayesian networks can take a long time and sometimes be unfeasible depending on the size of the network. The goal is to combine both score-based approach and constraint-based approach to form a hybrid approach, which subsequently reduces the runtime of pygobnilp.

University of BRISTOL