Probabilistic Reasoning

Unit # 15

Finding Node Ordering

- You might wonder where we could obtain the ordering required by K2.
- Such an ordering could possibly be obtained from domain knowledge such as a time ordering of the variables.
- For example, we might know that in patients, smoking precedes bronchitis and lung cancer and that each of these conditions precedes fatigue and a positive chest X-ray.

Algorithm without a Prior Ordering

- The following greedy search algorithm that does not require a time ordering.
- The search space is again the set of all DAGs containing the n variables.
- Following operations are allowed.
  1. If two nodes are not adjacent, add an edge between them in either direction.
  2. If two nodes are adjacent, remove the edge between them.
  3. If two nodes are adjacent, reverse the edge between them.

Algorithm without a Prior Ordering (Cont'd)

Problem: Find a DAG that approximates maximizing $score(\mathbb{G} : \mathsf{D})$.

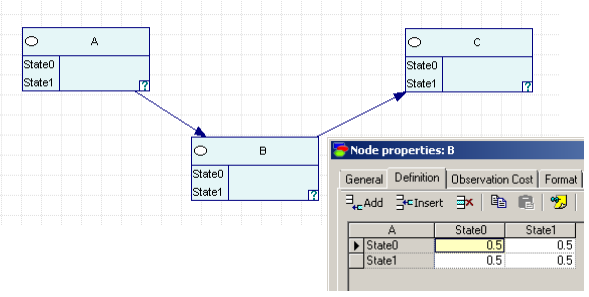
Inputs: A set V of n random variables; data D.

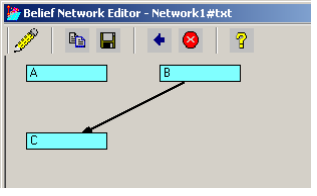
Outputs: A set of edges E in a DAG that approximates maximizing $score(\mathbb{G} : \mathsf{D})$.

```
void DAG_search (set_of_variables V, data D,
                 set_of_edges& E)
{
  E = ∅; G = (V, E);
  do
    if (any DAG in the neighborhood of our current DAG
    increases score(G : D))
      modify E according to the one that increases score(G : D) the most;
  while (some operation increases score(G : D));
}
```

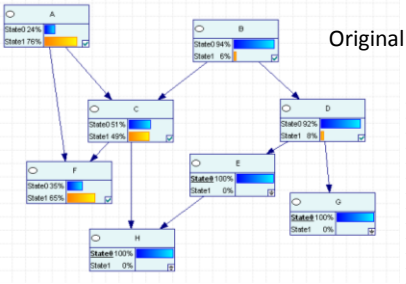
Structure Learning Example I

- Not a faithful DAG as due to the given conditional distribution A and B are independent.
- We generated 1000 records from the given network and then learn it using BN PowerConstructor.
- The network in the diagram below is learned by BN PowerConstructor.

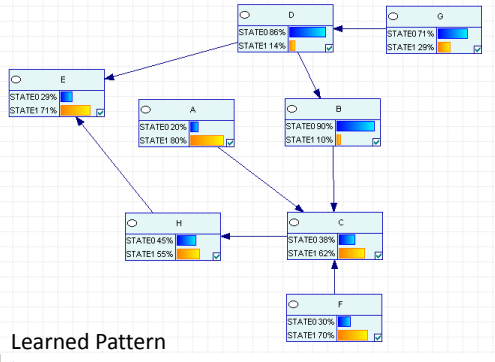


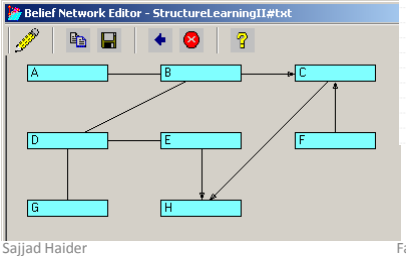


Structure Learning Example II



Original Model





Learned Pattern by BN PowerConstructor

Learned Model

Bayesian Information Criterion (BIC)

- The Bayesian information criterion (BIC) score is as follows:

$$BIC\,(\mathbb{G} : \mathsf{D}) = \ln\left(P(\mathsf{D}|\hat{\mathsf{P}}, \mathbb{G})\right) - \frac{d}{2}\ln m,$$

- where m is the number of data items and d is the dimension of the DAG model.
- The dimension is the number of parameters in the model.

Advantages of BIC

- The BIC score is intuitively appealing because it contains
  1. a term that shows how well the model predicts the data when the parameter set is equal to its ML value, and
  2. a term that punishes for model complexity.
- Another nice feature of the BIC is that it does not depend on the prior distribution of the parameters, which means there is no need to assess one.

Demo of BN PowerConstructor