Bayesian Networks without Tears
by Eugene Charniak

Discussion moderator: Isabelle Guyon

Summary

The paper provides a simple introduction to Bayesian networks, interpreting them as "causal" networks.

The basic concepts introduced are:

What do Bayesian networks represent?
- Bayesian networks are graphs (usually directed acyclic graphs, DAGs)
- The nodes represent random variables.
- The arrows represent dependencies between random variables and are sometimes interpreted as causal relationships.
- Each node X is associated with a conditional distribution P(X|parents), where parents are the nodes sending arrows to X. Root nodes X are associated with a prior P(X).
- The arrows are consistent with independence assumptions between the variables.

How are they designed?
- The P(X|parents) can be computed from real data. Even the structure of the network can to some extend be learned. However, the structure is most often set by hand using expert knowledge. The P(X|parents) are also often set "by hand". Probabilities in this last case are "beliefs" (hence the other name "belief networks").
- In practice, for discrete variables, we need to fill out a table of probability values for all values taken by the random variable of the node of consideration and all its parents. I.e. for binary variables, and a node with 2 parents, this makes 8 values. See the Genie software http://genie.sis.pitt.edu/.

What can they do for us?
- The entire network can be understood as a representation of the joint probability distributions of all the random variables of its nodes: $P(X_1, X_2, \ldots X_n)$, taking into account independence assumptions between the random variables. Generally the "chain rule" allows to compute $P(X_1, X_2, \ldots X_n)$ as $P(X_1|X_2, \ldots X_n)\ P(X_2|\ X_3, \ldots X_n) \ldots P(X_n)$. With the independence assumptions, it reduces to the product of $P(X_i|\text{parents})$, where hopefully the number of parents is small. Hence potentially important computational savings.
- The Bayesian network can allows us to compute conditional probabilities of the nodes, given that some of them have been observed: we update our "beliefs" in light of "evidence".

Can we read out variable dependencies?

- Two variables (or set of variables) are dependent if they are d-connected (that is they are not d-separated).
- For d-connectivity, we need a path linear or diverging not "blocked by evidence" or a converging path blocked by evidence.
- Here is the simple case of 3 variables. If e is in bold face, it means that it has been observed.

a and b are d-connected

$a \rightarrow e \rightarrow b$
$a \leftarrow e \rightarrow b$ } There is a "causal" path between a an b

$a \leftarrow e \leftarrow b$
$a \rightarrow \mathbf{e} \leftarrow b$ There is "evidence" making the variables correlated with each other

a and b are d-separated

$a \rightarrow \mathbf{e} \rightarrow b$
$a \leftarrow \mathbf{e} \rightarrow b$ } Evidence block the path
$a \leftarrow \mathbf{e} \leftarrow b$
$a \rightarrow e \leftarrow b$ The two variables are independent

How do we "evaluate" a network?

- The problem is NP-hard.
- Exact solutions exist for special cases of networks, e.g. polytrees (singly connected networks) and rely on the notion of d-separation to factor the problem into sub-graphs.
- Approximate solutions exist, using iterative procedures that randomly sample node values.

Criticism:

The representation of $P(X_1, X_2, \ldots X_n)$ as $\Pi_i P(X_i|\text{parents})$ is not unique. The author presents Bayesian networks as causal networks. Indeed, the direction of the arrows may be interpreted as causal dependencies. But, several Bayes nets with arrows NOT representing causal dependencies may compute correctly $P(X_1, X_2, \ldots X_n)$. This point is not made clear in the paper.

Suggested questions for discussion:

- Can all sets of independence assumptions be represented as a Bayesian network?
- What is the connection between causality inference and feature selection?