СТАТИСТИКА

Volokhovskyi V.Ye.
Student,
Kharkiv National University of Radio Electronics

BAYESIAN NETWORK

A Bayesian network is a probabilistic graphical model that represents a set of variables and their conditional dependencies through a directed acyclic graph. Bayesian networks are built from probability distributions and use the laws of probability for prediction and anomaly detection, for reasoning and diagnostics, decision making under uncertainty and time series prediction. Bayesian networks are represented using directed acyclic graphs whose nodes represent variables in the Bayesian sense: they may be observable quantities, unknown parameters or hypotheses. Edges show conditional dependencies between nodes; nodes that are not connected show variables that are conditionally independent of each other. Edges are added between nodes to indicate that one node directly influences the other. When an edge does not exist between two nodes, this does not mean that they are completely independent, these nodes can be connected through other nodes. They may however become dependent or independent depending on the evidence that is set on other nodes. Each node is associated with a probability function that takes, as input, a particular set of values for the node's parent variables, and gives the probability of the variable represented by the node.

The Bayesian network allows you to get answers to the following types of probabilistic requests:

- finding the probability of evidence,
- determination of a priori marginal probabilities,
- determination of a posteriori marginal probabilities, including: forecasting, or direct inference, – determining the likelihood of an event for the observed reasons; diagnosis, or the reverse conclusion (abduction), – determination of the probability of cause at the observed effects; inter causal inference or transduction, is the determination of the probability of one of the causes of an event, subject to the occurrence of one or several other causes of this event,
- calculation of the most likely explanation of the observed event,
- calculation of the a posteriori maximum.

Let G = (V,E) be a directed acyclic graph and let $X = (X_v)$, $v \in V$ be a set of random variables indexed by V [2, p. 490].

X is a Bayesian network with respect to G if its joint probability density function can be written as a product of the individual density functions, conditional on their parent variables:

$$p(x) = \prod_{v \in V} p\left(x_v \middle| x_{pa(v)}\right)$$

where pa(v) is the set of parents of v.

For any set of random variables, the probability of any member of a joint distribution can be calculated from conditional probabilities using the chain rule:

$$P(X_1 = x_1, \ldots, X_n = x_n) = \prod_{v=1}^{n} P(X_v = x_v \mid X_{v+1} = x_{v+1}, \ldots, X_n = x_n)$$

Local Markov property

X is a Bayesian network with respect to G if it satisfies the local Markov property which is defined in this way: each variable is conditionally independent of its non-descendants given its parent variables:

$$X_v \perp\!\!\!\perp X_{V \setminus de(v)} \mid X_{pa(v)}, \forall v \in V$$

where de(v) is the set of descendants and V \ de(v) is the set of non-descendants of v.

Bayesian network is a complete model for its variables and their relationships, so it can be used to answer probabilistic queries about them. The network can be used to update knowledge of the state of a subset of variables when other variables are observed. This process of computing the posterior distribution of variables given evidence is called probabilistic inference.

Bayesian network is specified by an expert and is then used to perform inference. In some problems the task of defining the network is too complex for people. In this case the network structure and the parameters of the local distributions must be learned from data.

An alternative method of structural learning uses optimization-based search. It requires a scoring function and a search strategy. A common scoring function is posterior probability of the structure given the training data. A local search strategy makes incremental changes aimed at improving the score of the structure. A global search algorithm like Markov chain Monte Carlo can avoid getting trapped in local minimum.

Fast method for exact Bayesian network learning is to cast the problem as an optimization problem, and solve it using integer programming. Acyclicity constraints are added to the integer program during solving in the form of cutting planes. Such method can handle problems with up to one hundred variables.

There is a method which can work with thousand variables. It implies working on the search space of the possible orderings, which is convenient as it is smaller than the space of network structures. Multiple orderings are then sampled and evaluated. This method is used when the number of variables is huge [1, p. 157].

Bayesian network give an opportunity not to use a lot of memory over exhaustive probability tables, if the dependencies in the joint distribution are sparse. The advantage of Bayesian networks is that it is easier for a people to understand direct dependencies and local distributions than complete joint distributions.

References:

1. Castillo E, Gutiérrez JM, Hadi AS. «Learning Bayesian Networks». Expert Systems and Probabilistic Network Models. Monographs in computer science. New York: Springer-Verlag. Pp. 481–528.
2. Pearl J, Russell S. «Bayesian Networks». In Arbib MA. Handbook of Brain Theory and Neural Networks. Cambridge, Massachusetts: Bradford Books. Pp. 157–160.