A Short Introduction to Bayesian Belief Network

# Bayesian Belief Network:

A BBN (Bayesian Belief Network) is a computational model which is based on graph probability theory. The structure of BBN is represented by Direct Acyclic Graph (DAG). Formally, a DAG is a pair *(N, A)* where *N* is the node set and *A* is the arc set. If there are two nodes *u* and *v* belonging to *N* and there is an arc going from *u* to *v* then *u* is termed as parent of *v* and *v* is called child of *u*. In terms of cause-effect relationship u is cause and v is effect. A node can be a parent for another node whereas the same node can be child for a different node. For example:

b

v

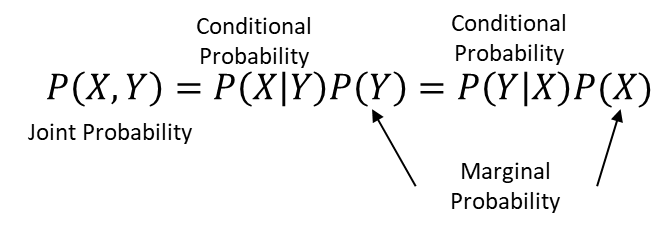
u

a

a and b are parents of u and u is child for a and b whereas u is a single parent for v. With respect to cause- effect relationship a and b are direct causes of u and u is directly causing v whereas a and b are indirectly responsible for occurrence of v.

## Bayes Theorem:

For two random variables X and Y the following equation holds true.



If the X and Y are independent of each other then

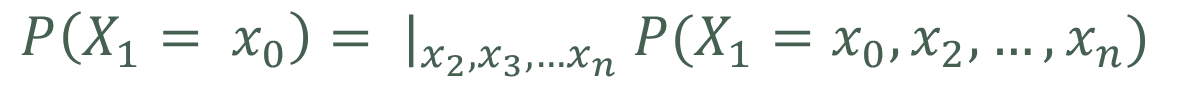


## Joint Probability:

Given X1, X2,………,Xn are features (nodes) in a BBN. Then Joint probability is defined as:

## Marginal Probability:

Given the joint probability, marginal probability of X1 = x0 is calculated as:



Where x2 ,x3,…xn are the set of values corresponding to X2, X3,…..,Xn

## Belief Propagation:

Let us understand with the help an example. We will consider a very basic network.

Strike, No Strike

T

K

A

Late, On time

Late, On time

Now above network says that Train strike influences Arvind’s and Karthik’s office timing. And the probabilities are distributed as below:



Now given we know Train Strike probability P(T) and conditional probabilities P(K|T) and P(A|T). we can calculate P(A) and P(K).

P(A= Y) = |T,K P(A=Y,T,K) = P(A=Y,K=Y,T=Y) + P(A=Y,K=N,T=Y) + P(A=Y,K=Y,T=N) + P(A=Y,K=N,T=N)

= P(T=Y)\*P(A=Y|T=Y)\*P(K=Y|T=Y) + P(T=Y)\*P(A=Y|T=Y)\*P(K=N|T=Y)

+ P(T=N)\*P(A=Y|T=N)\*P(K=Y|T=N) + P(T=N)\*P(A=Y|T=N)\*P(K=N|T=N)

= P(A=Y|T=Y)\*P(T=Y) + P(A=Y|T=N)\*P(T=N) …………………………………..(As, P(K=Y|T=Y) + P(K=N|T=Y) =1 )

= 0.7\*0.1 + 0.6\*0.9 = 0.61

Similarly, making the above formulation shorter.

P(K= Y) = P(K=Y|T=Y)\*P(T=Y) + P(K=Y|T=N)\*P(T=N)

= 0.6\*0.1 + 0.1\*0.9 = 0.15

Now, let us say we came to know that Arvind is late but we don’t know that if there is train strike. Can we know the probability that Karthik will be late given we know that Arvind is late? Let us see how the evidence that is Arvind is late propagates through the network.

Let us estimate the probability of train strike given we already know Arvind is late.

P(T=Y|A=Y) = P(A=Y|T=Y) \* P(T=Y)/P(A=Y) = 0.7\*0.1/0.61 = 0.12

This tells us that, if we Arvind is late then the probability that there is train strike is 0.12. We can use this updated belief of train strike to calculate the probability of Karthik being late.

P(K=Y) = P(K=Y,T=Y) + P(K=Y,T=N) = P(K=Y|T=Y)\*P(T=Y) + P(K=Y|T=N)\*P(T=N)

= 0.6\*0.12 + 0.1\*0.88 = 0.16

This gives slight increase in the probability of Karthik being late. So, the evidence that Arvind is late propagates through the network and changes the belief of train strike and Karthik being late.

Prior Belief

Evidence

**Posterior Belief**

# Constructing Discrete Bayesian Belief Network:

BBN can be constructed using only continuous variables, only categorical variables or mix of variables. We will discuss discrete BBN which is constructed using categorical variables only. There are two major components of constructing a Bayesian belief network. Structure Learning and Parameter Learning.

## Structure Learning:

Structure leaning is the basis of Bayesian Belief Network analysis. Effectiveness of the solution depends on the optimality of the learnt structure. We can use the following approaches:

1. One can use expert view and cerate the structure based on the domain knowledge.
2. Using machine learning algorithm, we can learn local optimal structure. Please note global optimal structure is NP -hard problem. There are many algorithms to learn the structure like, K2, hill climbing and tabu etc. Please refer to “bnlearn” package in R and the following link for Python packages. <http://pgmpy.chrisittner.de/>
3. Combination of both. Use machine learning technique to build the model and with the reduced set of explanatory variables use expert view or machine learning technique to create the structure. This is could be a quick and effective way.

## Parameter Learning:

Another major component of BBN is CPT (Conditional Probability Table). Since each node in the structure is a random variable it can take multiple values/ states. Each state will have some probability of occurrence. We call these probabilities as Beliefs. Also, each node is connected to other nodes in the network. As per the structure we learn the conditional probability of each state of a node. Tabular form of all such probabilities are called CPT.

# Conclusion:

The document is scoped to the bare minimum concepts which are required to construct a discrete BBN and understand its various components. Structure learning and Parameter learning are the two major components which are required to construct a BBN whereas the concepts of Bayes theorem, joint and marginal probability work as base for the Network. Propagation of evidence required to understand the functionality of BBN.

BBN can be used like any other machine learning however, it works the best where there are interdependencies among the predictors and number of predictors are less. The best part of the BBN is its intuitive way of explaining the drivers of an evidence.