

Bayesian Networks for Causal Analysis

Fei Wang and John Amrhein, McDougall Scientific Ltd.

ABSTRACT

Bayesian Networks (BN) are a type of graphical model that represent relationships between random variables. The networks can be very complex with many layers of interactions. Graphical models become BNs when the relationships are probabilistic and uni-directional. Building BNs for causal analyses is a natural and reliable way of expressing (and confirming or refuting) our belief and knowledge about cause and effects. In addition, BNs can be easily reconfigured with minor modifications to facilitate our understanding of probabilistic mechanisms. This paper describes the construction of BNs for causal analyses and how to infer causal structures from observational and interventional data. The paper includes applications of causal BNs for classification using the HPBN Classifier node in Enterprise Miner. Visualization, inferences, and scenario analyses for the examples are discussed.

BAYESIAN CONCEPT OVERVIEW

BAYES' THEOREM

Bayes' theorem (Bayes' law or Bayes' rule) describes the probability of an event based on prior knowledge of conditions that might be related to the event. It can be stated mathematically as the following equation:

$$P(A|B) = \frac{P(A,B)}{P(B)} \quad (1)$$

where A and B are observed events, P(B) is the probability of observing B independently, P(A,B) is a joint probability of A and B which represents the likelihood of event A and B occurring together, and P(A|B) is a conditional probability which represents the likelihood of event A occurring given that B is observed.

An algebraic inversion of (1) can be written as:

$$P(A, B) = P(A|B)P(B) = P(B|A)P(A) \quad (1')$$

There are other equivalent statements of Bayes' theorem, each being useful in a different context.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad \text{or} \quad P(A|B) = P(A) \frac{P(B|A)}{P(B)} \quad (2)$$

$$P(A|B) = \frac{P(B|A)P(A)}{\sum_i P(B|A_i)P(A_i)} \quad \text{or} \quad P(A|B) = P(A) \frac{P(B|A)}{\sum_i P(B|A_i)P(A_i)} \quad (3)$$

Equation (1) gives rise to (2). Equation (3) stems from (2) by substituting the unconditional (marginal) probability of event B with the sum of its partitions over all possible outcomes of event A.

INTERPRETATION

The interpretation of Bayes' theorem depends on the interpretation of probability. In Bayesian interpretation, probability measures a degree of belief; while in frequentist interpretation, probability measures a mathematical frequency of outcomes.

Bayesians interpret Bayes' theorem as: the updated (posterior) belief of event A given that event B has been observed, P(A|B), is dependent upon the prior belief of event A, P(A), the likelihood of event B given that event A has occurred, P(B|A), and the prior belief of event B, P(B).

Frequentists, on the other hand, interpret Bayes' theorem as: over repeated sampling, P(A|B) is the proportion of outcomes with A out of all outcomes with B.

ADVANTAGE VS. DISADVANTAGE

There are a few advantages of using Bayesian concepts and analytical methods: 1) Bayesian methods can incorporate prior information, knowledge, and subject matter expertise without corresponding data; 2) Bayesian methods enable the study of a cause-effect interpretations rather than just correlations.

However, the sometimes-subjective nature of priors is a disadvantage of Bayesian methods. The posterior distribution may be heavily influenced by a subjective prior in some situations. One can use non-informative priors to mitigate the impact of subjectivity on the posterior estimates.

BAYESIAN NETWORK

DEFINITIONS AND PROPERTIES

A Bayesian Network (BN) is a representation of a joint probability distribution of a set of random variables with probabilistic dependencies. It is a class of graphic models that consist of two parts, $\langle \mathbf{G}, \mathbf{P} \rangle$:

- \mathbf{G} is a directed acyclic graph (DAG) made up of nodes corresponding to random variables, \mathbf{X} in \mathbf{U} , and arcs (edges, links, or connectors) representing conditional dependencies between random variables
- \mathbf{P} is a set of conditional probability distributions, one for each node conditional on its parents.

Figure 1 shows an example of a DAG. Letters A – G represent nodes. The graph is directed because

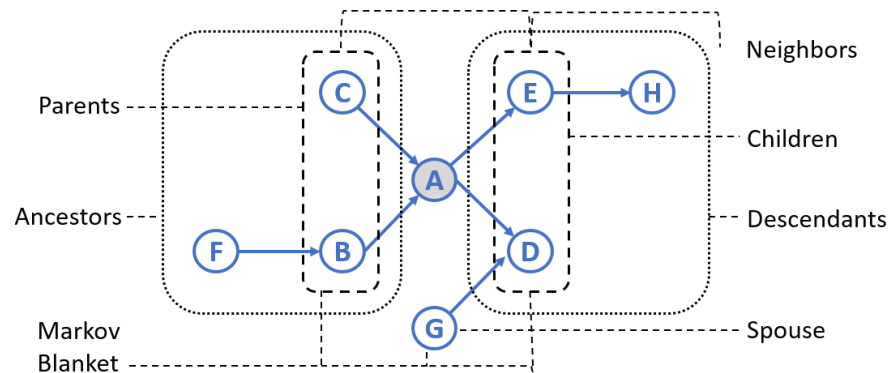


Figure 1 Parents, Ancestors, Neighbors, Spouse, Children, Descendants, and Markov Blanket of Node A in a DAG

each arc between two nodes is uniquely directed, and is acyclic because no cycles or loops (e.g. $A \rightarrow B \rightarrow C \rightarrow A$) exist. A node from which a directed edge starts is called the parent of the node to which the edge is directed; a node on which a directed edge ends is called the child of the node from which it comes. For example, in Figure 1, nodes B and C are parents of node A, and node A is the common child of nodes B and C. In the example for node A, nodes B, C, and F are ancestors of node A; nodes D, E, and H are descendants of node A. Parents and children of a node are the neighbors of the node. Spouses are nodes sharing the same child, such as A and G. Parents, children, and spouses of a node constitute the Markov blanket of the node.

The component \mathbf{P} of a BN is a set of conditional probability distributions, one for each node conditional on its parents. It follows the Markov property of BN: *each node is conditionally independent of its ancestors given its parents*. One can claim that changes in the belief of a node have no impact on the belief of its ancestors given its parents.

The Markov property of BN enables the presentation of the joint probability distribution of random variables in \mathbf{G} decomposed into a product of conditional probability distributions. For Figure 1, we have

$$P(A, B, C, D, E, F, G, H) = P(A|B, C)P(B|F)P(C)P(D|A, G)P(E|A)P(F)P(G)P(H|E).$$

It is easy to see that, given the Markov blanket, a node is independent from the rest of nodes. Therefore, the probability distribution of a node is completely determined by its Markov blanket.

Once a BN is learnt from data, one can investigate the effects of a new piece of evidence $E (E \subseteq U)$ in the distribution of U using the knowledge encoded in the BN, that is, to investigate the posterior probability $P(Q|E, G, P)$, where $Q \cup E = U$. The posterior probability can be interpreted in terms of the changes in one's beliefs according to the observed new evidence.

BN LEARNING IN SAS®

In many practical settings, the BN is unknown and needs to be learnt from data. The task of fitting a BN is known as BN learning. Given training data and prior information, such as expert knowledge about possible casual relationships, BN learning performed in two steps; structure learning and parameter learning.

Structure Learning

The first step consists in identifying the graph structure of the BN; i.e. how the variables relate to each other as parents, children, neighbors, and spouses. Ideally, the final learnt structure should result in a joint posterior distribution as close as possible to the correct one in the probability space.

The HPBNET procedure deployed in the HPBN node in SAS® Enterprise Miner™ (EM) is a high-performance procedure which can learn different types of BN structures from an input data set. The supported types of network structures are naïve Bayesian (NB), tree-augmented naïve (TAN), Bayesian network-augmented naïve (BAN), parent-child Bayesian network (PC), and Markov blanket (MB). Refer to the reference by Liu et al (2017) for descriptions of each type. Briefly, NB, TAN, and BAN all have one target variable that is a parent to each input variable. NB is similar to a typical linear regression in that the only dependency is between a target variable and the inputs. The inputs themselves are conditionally independent of each other given the target. A TAN extends the NB by allowing a tree structure among the inputs, and a BAN loosens the structure among the inputs to be a BN itself. A PC BN allows inputs to be parents of the target, and the MB is the least restrictive as shown in Figure 1.

Based on the specified BN type, PROC HPBNET uses different algorithms to learn the relationships between the variables, i.e. to define the structure of the graph.

Two categories of algorithms are used in PROC HPBNET:

- The score-based approach uses the Bayesian information criterion (BIC) to measure the goodness of fit of a structure based on the training data, and to find a structure with the highest BIC score. The BIC is a penalized likelihood score. For discrete data, it is defined as

$$BIC = N \sum_{i=1}^n \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} p(\pi_{ij}) p(X_i = x_{ik} | \pi_{ij}) \ln p(X_i = x_{ik} | \pi_{ij}) - \frac{M}{2} \ln N$$

where N is the number of observations in the training data, n is the number of variables (nodes), X_i is a random variable, r_i is the number of levels for X_i , x_{ik} is the k th value of X_i , q_i is the number of value combinations of X_i 's parents, π_{ij} is the j th value of X_i 's parents, and $M = \sum_{i=1}^n (r_i - 1) \times q_i$ is the number of parameters for the probability distributions.

- The constraint-based approach uses independence tests to determine the edges and the directions. In some cases when the direction of the edge cannot be oriented, PROC HPBNET uses the BIC score to determine directions of the edges.

When there are many input variables, the number of variable combinations is exponential and structure learning is time consuming. Therefore, variable selection is needed.

Because the probability distribution of a node is completely determined by its Markov blanket, the Markov blanket can be used for variable selection. PROC HPBNET supports two types of variable selection: independence tests and conditional independence tests. Independence tests, such as a chi-square test for discrete data, are performed between each input variables and the target variable. A significant result

of the independence test indicates an edge between the input variable and the target variable. While conditional independence tests are performed between each input variable and the target variable given any subset of other input variables.

Take the MB BN structure type for example. PROC HPBNET learns the parents of the target variables first by applying independence tests to determine the edges. The edges are then oriented by using conditional tests and the BIC score. Then PROC HPBNET learns the parents of the input variable which has the highest BIC score with the target by using the same method, and continues to learn the parents of the input variable which has the second highest BIC score with the target.

Parameter Learning

Parameter learning determines the probability distribution of each node in a BN. Once the structure of the network has been learnt from the data, estimating the parameters of the global distribution is greatly simplified by the application of the Markov property.

One can also classify the value of a target variable using the posterior probability distribution and an observation of other variables. The predicted class is the one that has the largest posterior probability. That is, for an observation of the input variables, $(X_1 = x_1, X_2 = x_2, \dots, X_{n-1} = x_{n-1})$, the predicted class of target T is c satisfying

$$\max_c P(T = c | X_1 = x_1, X_2 = x_2, \dots, X_{n-1} = x_{n-1})$$

where

$$\begin{aligned} P(T = c | X_1 = x_1, X_2 = x_2, \dots, X_{n-1} = x_{n-1}) &\propto P(T = c, X_1 = x_1, X_2 = x_2, \dots, X_{n-1} = x_{n-1}) \\ &= \prod_{i=1}^{n-1} P(X_i = x_i | \text{parents}(X_i)) P(T = c | \text{parents}(T)). \end{aligned}$$

For example, Figure 2 shows a learnt BN from a dataset in which all variables have a value of either “Yes”

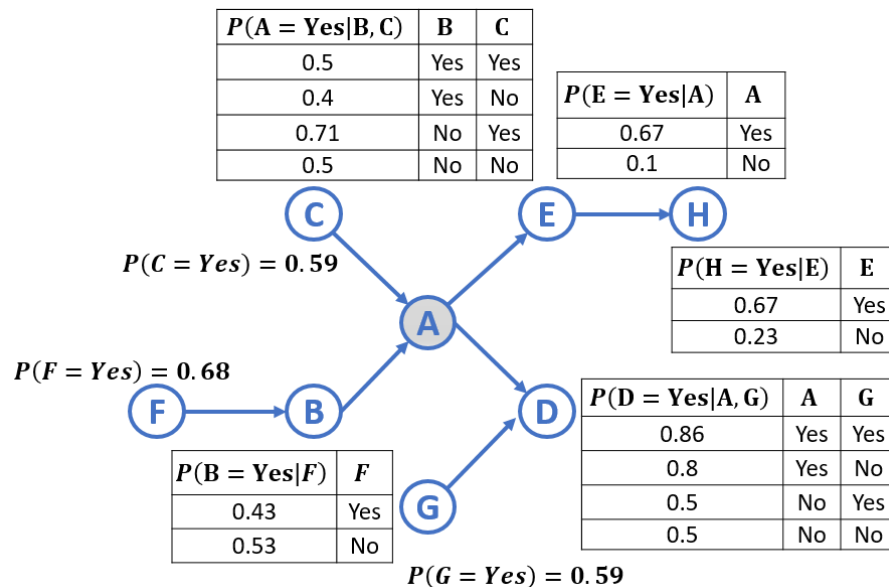


Figure 2 A Bayesian Network

or “No”. Suppose we observe an event $(B=\text{Yes}, C=\text{Yes}, D=\text{Yes}, E=\text{No}, F=\text{Yes}, G=\text{No}, H=\text{Yes})$, we want to predict the value of A. Based on the learnt BN in Figure 2, we have

$$\begin{aligned} P(A = Y | B = Y, C = Y, D = Y, E = N, F = Y, G = N, H = Y) \\ \propto P(A = Y, B = Y, C = Y, D = Y, E = N, F = Y, G = N, H = Y) \end{aligned}$$

$$\begin{aligned}
&= P(A = Y|B = Y, C = Y)P(B = Y|F = Y)P(C = Y)P(D = Y|A = Y, G = N)P(E = N|A = Y)P(F = Y)P(G = N)P(H = Y|E = N) \\
&= 0.5 \times 0.43 \times 0.59 \times 0.8 \times (1 - 0.67) \times 0.68 \times (1 - 0.59) \times 0.23 \\
&= 0.002;
\end{aligned}$$

And

$$\begin{aligned}
&P(A = N|B = Y, C = Y, D = Y, E = N, F = Y, G = N, H = Y) \\
&\propto P(A = N, B = Y, C = Y, D = Y, E = N, F = Y, G = N, H = Y) \\
&= P(A = N|B = Y, C = Y)P(B = Y|F = Y)P(C = Y)P(D = Y|A = N, G = N)P(E = N|A = N)P(F = Y)P(G = N)P(H = Y|E = N) \\
&= (1 - 0.5) \times 0.43 \times 0.59 \times 0.5 \times (1 - 0.1) \times 0.68 \times (1 - 0.59) \times 0.23 \\
&= 0.004.
\end{aligned}$$

Thus, if an event with (B=Yes, C=Yes, D=Yes, E=No, F=Yes, G=No, H=Yes) is observed, A will most likely happen with a value of No

Note that, the normal constant $P(B = Y, C = Y, D = Y, E = N, F = Y, G = N, H = Y)$ is eliminated from the calculation of the two conditional probabilities of A given other nodes. The exact conditional probability can be obtained by dividing the normal constant which can be easily calculated from the learnt BN.

CAUSAL INFERENCE

Suppose we have learnt a BN with structure \mathbf{G} and conditional probability distribution \mathbf{P} .

When a BN is given a causal interpretation, the arcs describe cause-and-effect relationships instead of probabilistic dependencies. For instance, according to Pearl's *Causality*, in a causal BN, the parents of each node are its direct causes, ancestors other than parents are its indirect causes, children of each node are its direct effects, and descendants other than children are its indirect effects.

To evaluate the conditional probability of known causes given their effects and vice versa, three additional assumptions are needed:

- Each node is conditionally independent of its indirect causes, given its direct causes. This assumption is called the causal Markov property, and is the causal interpretation of the Markov property.
- There must exist a network structure which represents the true dependence structure of \mathbf{G} . This assumption is usually referred as the faithfulness assumption.
- There must be no latent variables which are unobserved variables influencing the variables in the network, acting as confounding factors.

The third assumption is a deduction of the first two. The presence of latent variables may induce wrong correlations between observed variables. If an arc is wrongly added, the causal Markov property might be violated. And, the presence of latent variables violates the assumption of faithfulness.

In this setting, the posterior probability is interpreted as measures of the effects of interventions (denoted as \mathbf{I} , and $\mathbf{I} \subseteq \mathbf{U}$) on the causal structure. The intervention represents an action whose only effect is to fix the values of the variables in \mathbf{I} to specific values, such as $\mathbf{I} = (X_{i1} = x_{i1}, X_{i2} = x_{i2}, \dots, X_{ik} = x_{ik})$, where $i1, \dots, ik \in \{1, 2, \dots, n\}$. Then the conditional probability $P(\mathbf{Q}|\mathbf{I}, \mathbf{G}, \mathbf{P})$ studies the effect of interventions \mathbf{I} .

According to the causal Markov property of the BN, $P(\mathbf{Q}|\mathbf{I}, \mathbf{G}, \mathbf{P})$ can be calculated as follows:

$$P(\mathbf{Q}|\mathbf{I}, \mathbf{G}, \mathbf{P}) = P(\mathbf{I}) \times \prod P(\mathbf{Q}|\text{Parents}(\mathbf{G}, \mathbf{P})),$$

where $\text{Parents}(\mathbf{G}, \mathbf{P})$ is the set of parents for variables in \mathbf{Q} .

In discrete case of data, $\prod P(\mathbf{Q}|\text{Parents}(\mathbf{G}, \mathbf{P}))$ can be obtained by using the relative frequency of \mathbf{Q} conditional on $\text{Parents}(\mathbf{G}, \mathbf{P})$.

EXAMPLE

The example is a small synthetic data set from Lauritzen and Spiegelhalter (1988) about lung diseases (tuberculosis, lung cancer, and bronchitis), in which the example is motivated as follows:

“Shortness-of-breath (dyspnea) may be due to tuberculosis, lung cancer or bronchitis, or none of them, or more than one of them. A recent visit to Asia increases the chances of tuberculosis, while smoking is known to be a risk factor for both lung cancer and bronchitis. The results of a single chest X-ray do not discriminate between lung cancer and tuberculosis, as neither does the presence or absence of dyspnea.”

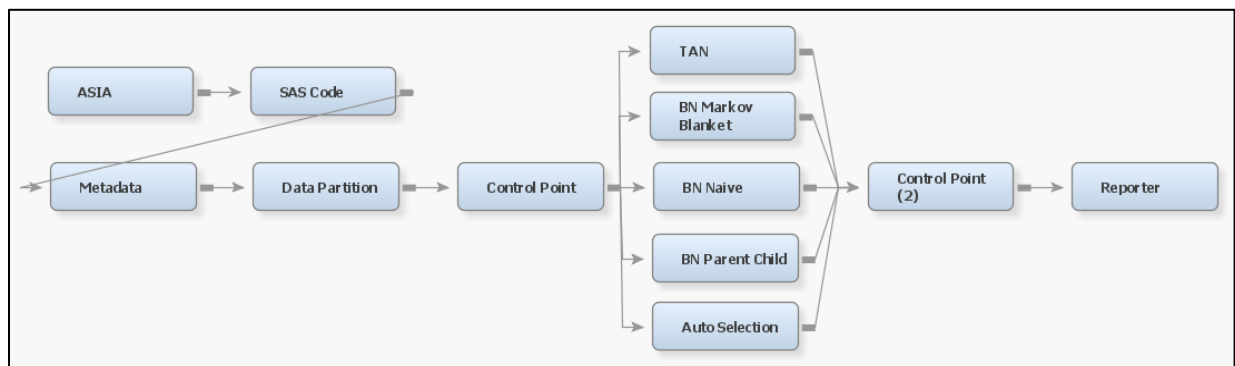
The data set contains 5000 observations and 8 binary character variables (Table 1) whose values are either yes or no. Suppose that we want to determine the dependencies between a set of inputs, dyspnea, bronchitis, a visit to Asia, smoking, and a positive x-ray, and the verified presence of either tuberculosis or lung cancer.

Table 1. Details of the Asia Data Set

Variable	Label	Description	Frequency “Yes”
Dyspnea	Dyspnea	Presence of dyspnea?	2350
Tuber	Tuberculosis	Presence of tuberculosis?	44
Cancer	Lung Cancer	Presence of lung cancer?	330
Bronch	Bronchitis	Presence of bronchitis?	2549
Visit	Visit to Asia	Visit to Asia?	42
Smoke	Smoking	Is a smoker?	2515
XrayPos	Chest X-ray	Positive chest X-ray?	569
TubOrCan	Tuberculosis or Lung Cancer	Presence of tuberculosis or lung cancer?	370

Display 1 shows the EM process diagram. The SAS Code node is used to rename variables and assign labels. The Metadata node sets *TubOrCan* as the target, and the Data Partition node splits the data 3:1 into training and validation data respectively. We configured four structure types, TAN, MB, Naïve, and PC. The fifth HPBN node, Auto Selection, is configured for EM to auto-select the best fitting structure based on criterion such as misclassification rate.

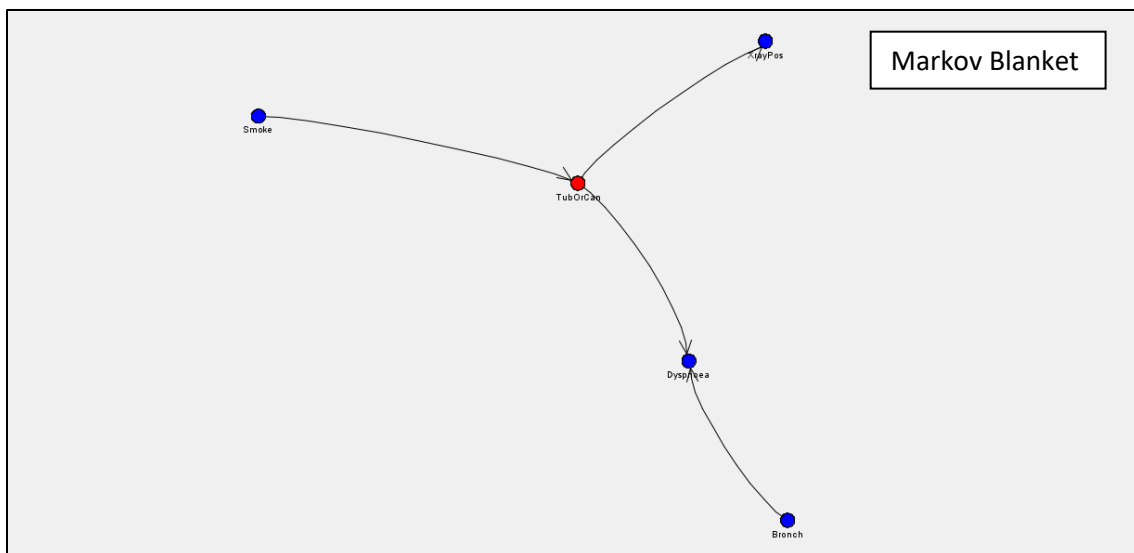
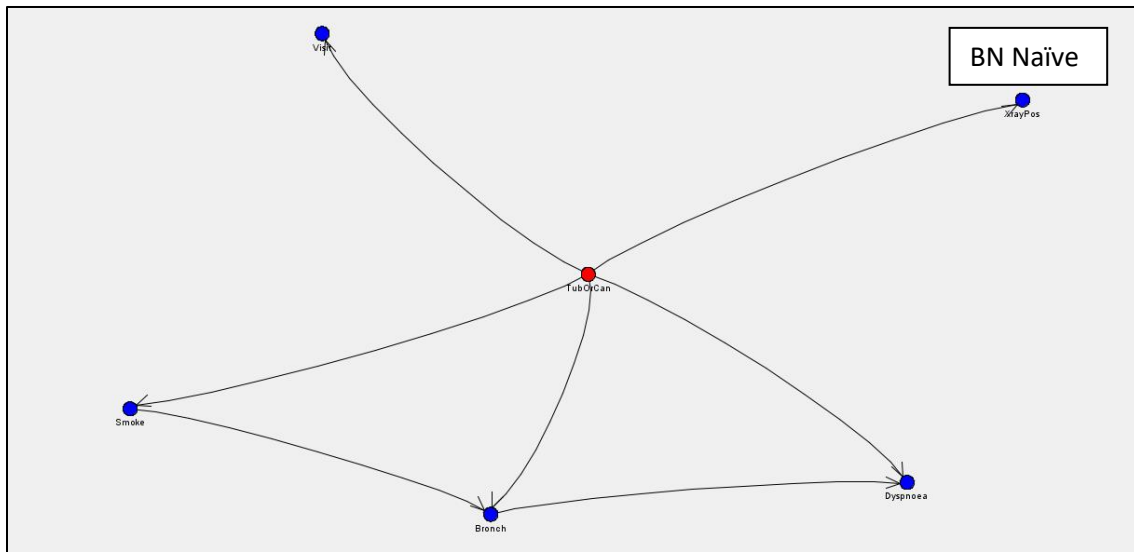
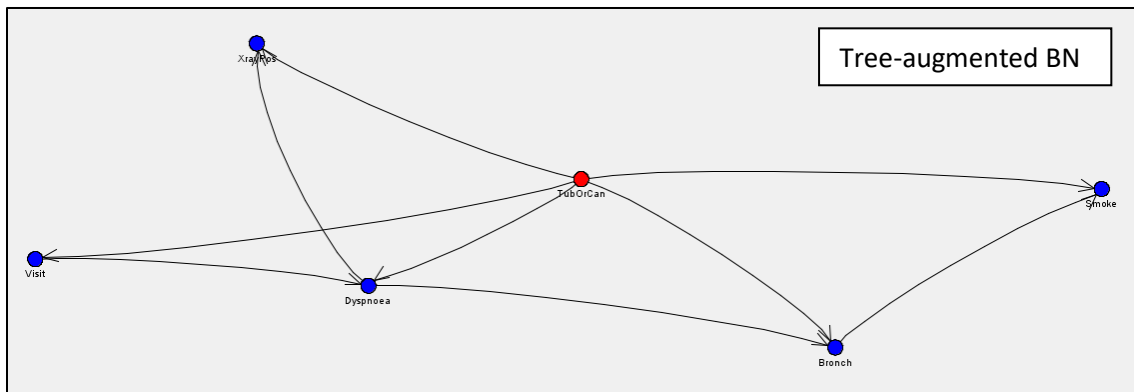
Display 1. Bayesian Network Diagram Based on Asia Data Set

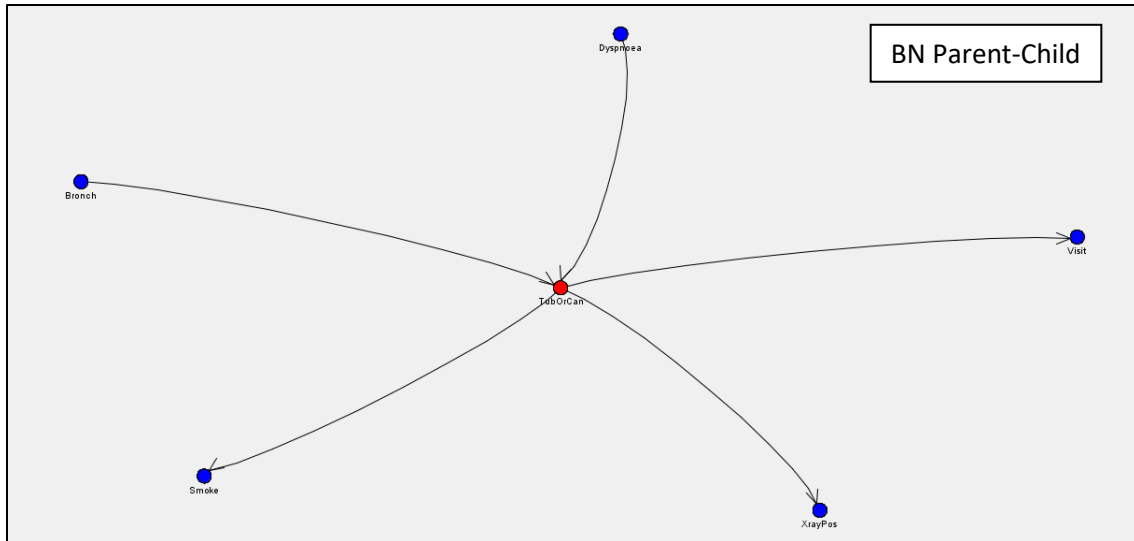


We configured the HPBN nodes to not use the tuberculosis and cancer variables. We also turned off variable pre-screening and selection because we wanted to explore the dependencies among the other

inputs. Therefore, our possible models are reduced to the various BN structure types and the edges connecting the nodes. Display 2 shows the final BN structure for each fitted structure type.

Display 2. Four BN Structures for the Asia Data





The learnt structures differ because of the restrictions placed on the connections within each type. The HPBN node, which is configured for auto-selection, chose the BN Parent-Child structure. The target variable is always at the center and is shown as red. The inputs are blue and the direction of the arcs are from parent to child. Posterior probabilities are given as $P(\text{child} | \text{parent})$. You must consider the temporal relationship between the parent and child, which HPBN will not know, to properly state probable cause for an effect.

For example, the *Smoke* node is a child of *TubOrCan*. The posterior conditional probability of smoking (cause) given tuberculosis or cancer (effect) is 0.846 (Output 1). Because we know that smoking pre-dates disease onset, **if this network completely represented all factors influencing disease onset**, then we could conclude, with very high probability, that smoking causes tuberculosis or cancer.

Output 1. BN-PC Posterior Probabilities of Smoking Given Tuberculosis or Cancer

Parent Node	Parent Condition	Child Node	Child Condition	Probability
TubOrCan	YES	Smoke	NO	0.154122
TubOrCan	YES	Smoke	YES	0.845878
TubOrCan	NO	Smoke	NO	0.524755
TubOrCan	NO	Smoke	YES	0.475245

The two parents of *TubOrCan*, bronchitis and dyspnea, may be symptoms of tuberculosis or cancer. Therefore, the posterior conditional probabilities could be used as diagnostic aids.

Output 2. BN-PC Posterior Probabilities of Tuberculosis or Cancer Given Dyspnea and Bronchitis

Parent Node	Parent Condition	Child Node	Child Condition	Probability
Bronch	NO	TubOrCan	YES	0.018205
Dyspnoea	NO	TubOrCan	YES	0.018205
Bronch	NO	TubOrCan	NO	0.981795
Dyspnoea	NO	TubOrCan	NO	0.981795
Bronch	NO	TubOrCan	YES	0.314961
Dyspnoea	YES	TubOrCan	YES	0.314961
Bronch	NO	TubOrCan	NO	0.685039
Dyspnoea	YES	TubOrCan	NO	0.685039
Bronch	YES	TubOrCan	YES	0.063415
Dyspnoea	NO	TubOrCan	YES	0.063415
Bronch	YES	TubOrCan	NO	0.936585
Dyspnoea	NO	TubOrCan	NO	0.936585
Bronch	YES	TubOrCan	YES	0.097333
Dyspnoea	YES	TubOrCan	YES	0.097333
Bronch	YES	TubOrCan	NO	0.902667
Dyspnoea	YES	TubOrCan	NO	0.902667

For example, from the table in Output 2, a patient who presents with shortness-of-breath but without bronchitis has a 0.315 probability of having tuberculosis or cancer. Stated another way, tuberculosis or cancer causes shortness of breath with probability 0.315.

The fitted BN-PC structure differs from the “true” structure in Lauritzen and Spiegelhalter (1988). The Markov blanket structure replicates their structure except that we omitted the individual indicators *Tuber* and *Cancer*, which are parents of

TubOrCan. They also show a directed arc from *Visit* into *Tuber*. Rather than choosing the BN-PC structure, which is strictly data-driven, we could choose the Markov blanket structure based on subject matter expertise that states that bronchitis is a probable cause of shortness of breath. That is, bronchitis should be a parent of dyspnea rather than a spouse, as in the PC structure.

The Markov blanket posterior probabilities for bronchitis, smoking, tuberculosis-or-cancer, and dyspnea, are shown in Output 3.

Output 3. Markov Blanket BN Posterior Probabilities

Child Node	Child Condition	Probability
Bronch	NO	0.491602
Bronch	YES	0.508398

Child Node	Child Condition	Probability
Smoke	NO	0.497201
Smoke	YES	0.502799

Parent Node	Parent Condition	Child Node	Child Condition	Probability
Smoke	NO	TubOrCan	YES	0.023044
Smoke	NO	TubOrCan	NO	0.976956
Smoke	YES	TubOrCan	YES	0.125066
Smoke	YES	TubOrCan	NO	0.874934

Parent Node	Parent Condition	Child Node	Child Condition	Probability
TubOrCan	YES	Dyspnoea	NO	0.266055
Bronch	NO	Dyspnoea	NO	0.266055
TubOrCan	YES	Dyspnoea	YES	0.733945
Bronch	NO	Dyspnoea	YES	0.733945
TubOrCan	YES	Dyspnoea	NO	0.151163
Bronch	YES	Dyspnoea	NO	0.151163
TubOrCan	YES	Dyspnoea	YES	0.848837
Bronch	YES	Dyspnoea	YES	0.848837
TubOrCan	NO	Dyspnoea	NO	0.899885
Bronch	NO	Dyspnoea	NO	0.899885
TubOrCan	NO	Dyspnoea	YES	0.100115
Bronch	NO	Dyspnoea	YES	0.100115
TubOrCan	NO	Dyspnoea	NO	0.220944
Bronch	YES	Dyspnoea	NO	0.220944
TubOrCan	NO	Dyspnoea	YES	0.779056
Bronch	YES	Dyspnoea	YES	0.779056

You can conduct scenario analyses to assess the impact of an intervention by fixing the value of the corresponding node. For example, to determine if bronchitis is a cause of dyspnea, the variable *Bronch* can be set to “yes”; i.e. $P(\text{Bronch} = \text{yes}) = 1$. To assess causality for dyspnea, we calculate: $P(\text{Dyspnea} = \text{yes} \mid \text{Bronch} = \text{yes})$. Because *Dyspnea* also has *TubOrCan* as a parent, we must calculate this probability over both values of *TubOrCan*. I.e.

$P(\text{Dyspnea} = \text{yes} \mid \text{Bronch} = \text{yes}, \text{TubOrCan} = \text{yes}) \times P(\text{TubOrCan} = \text{yes}) + P(\text{Dyspnea} = \text{yes} \mid \text{Bronch} = \text{yes}, \text{TubOrCan} = \text{no}) \times P(\text{TubOrCan} = \text{no})$. Because *TubOrCan* also has a parent, *Smoke*, its probabilities must be calculated by summing over values of *Smoke*. The complete calculation is:

$$P(\text{Dyspnea} = \text{yes} \mid \text{Bronch} = \text{yes}) = .85 \times (.02 \times .5 + .13 \times .5) + .78 \times (.98 \times .5 + .87 \times .5) = 0.79.$$

Similarly, we calculate the probability of *Dyspnea* in the absence of bronchitis;

$$.73 \times (.02 \times .5 + .13 \times .5) + .10 \times (.98 \times .5 + .87 \times .5) = 0.15.$$

To determine the effect of bronchitis on the shortness-of-breath, we can calculate the difference, $0.79 - 0.15 = 0.64$, or risk ratio, $0.79/0.15 = 5.27$. We can conclude that Bronchitis is a positive cause of *Dyspnea*. Of course, because the probability tables are stored as data sets, you can automate calculations of this sort if they are known to be of interest in advance.

It is also possible to score a data set of observations to obtain predictions of interest. For example, we created two observations with input values as shown in Output 4. The difference between the two is the value of smoking; one is set to “yes” and the other “no”. The predicted probability for the outcome,

TubOrCan, is stored in the variable *P_TubOrCanYES*. Because HPBN is a classifier, EM will automatically assign each observation to an outcome class with the higher predicted probability.

Output 4 Two Observations Scored Using the BN Markov Blanket

Obs	Visit	Bronch	Dyspnoea	Tub		Cancer	Smoke	Tuberc	Xray Pos	_WARN_	P_Tub	P_Tub	I_
				OrCan	OrCan						OrCan	OrCan	
1	No	Yes	Yes				Yes		Yes		0.68621	0.31379	YES
2	No	Yes	Yes				No		Yes		0.39447	0.60553	NO

CONCLUSION

This paper describes Bayesian networks (BN), the construction of BNs in SAS®, and how to use BNs for causal inference. An example based on the *Asia* data set is given by an implementation in SAS® Enterprise Miner using the HPBN classifier node.

In cases with continuous variables, SAS® Enterprise Miner bins continuous variables into equal-width levels and treats them as categorical variables. The default number of levels is 5, and users can manually specify the number of levels.

BN learning methods described in this paper are assumed to be based on observational and experimental data where observations in a sample are assumed to be collected under the same general conditions. In cases with interventional data, values of specific variables (interventions) are set by an external intervention. BN structures can be learnt from a structure in which the intervention variables are included and all other variables are depended on them. It can be done by adding specific arcs from intervention variables into other variables, which could be realized in the future release of HPBNET node in SAS® Enterprise Miner.

Real-world variables or factors comprise complex systems which may evolve over time. The set of values of the variables or factors are a time-stamped state and, at any point in time, there is a prior state, transition, and current state. Dynamic Bayesian networks can be used to model dependencies between factors of interests arising from such time series. Examples include system reliability models and models of operational risk in finance.

REFERENCES

- Amrhein, J., Wang, F. 2018. "Bayesian Concepts: An Introduction". Proceedings of SAS Global Forum 2018, Paper 1863-2018. Denver, Colorado. SAS Institute, Cary NC.
- Liu, Y., Shi, W., and Czika, W. 2017. "Building Bayesian Network Classifiers Using the HPBNET Procedure." Proceedings of SAS Global Forum 2017, Paper 474-2017. Orlando, FL. SAS Institute, Cary NC.
- Pearl, J. 2009. *Causality: Models, Reasoning, and Inference*. 2nd ed. Cambridge: Cambridge University Press
- Pearl, J., Glymour, M., Jewell, N. 2016. *Causal Inference in Statistics. A Primer*. Wiley
- Lautitzen, S.L. and Spiegelhalter, D.J. 1988. "A Local Computations with Probabilities on Graphical Structures and Their Applications to Expert Systems." *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 50, No. 2, 157-224.

ACKNOWLEDGMENTS

We would like to thank the Statistics Team in McDougall Scientific Ltd. for their support and advice.

RECOMMENDED READING

- SAS® Enterprise Miner™ 14.3 Guide

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Fei Wang
McDougall Scientific Ltd.
fwang@mcdougallscientific.com
<http://www.mcdougallscientific.com>