We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS

BOOK CITATION INDEX

INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Learning parameters and structure of Bayesian networks using an Implicit framework

Hanen Ben Hassen*, Lobna Bouchaala*,
Afif Masmoudi** and Ahmed Rebai*
*Unit of Bioinformatics and Biostatistics, Centre of Biotechnology of Sfax, Tunisia
** Laboratory of Probability and Statistics, Faculty of Science of Sfax, Tunisia

1. Introduction

A large amount of work has been done in the last ten years on learning parameters and structure in Bayesian networks (BNs) (see for example Neapolitan, 2005). Within the classical Bayesian framework, learning parameters in BNs is based on priors; a prior distribution of the parameters (prior conditional probabilities) is chosen and a posterior distribution is then derived given the data and priors, using different estimations procedures (for example Maximum a posteriori (MAP) or Maximum likelihood (ML),...). The Achille's heal of the Bayesian framework resides in the choice of priors. Defenders of the Bayesian approach argue that using priors is, in contrary, the strength of this approach because it is an intuitive way to take into account the available or experts knowledge on the problem. On the other side, contradictors of the Bayesian paradigm have claimed that the choice of a prior is meaningless and unjustified in the absence of prior knowledge and that different choices of priors may not lead to the same estimators. In this context, the choice of priors for learning parameters in BNs has remained problematic and a controversial issue, although some studies have claimed that the sensitivity to priors is weak when the learning database is large.

Another important issue in parameter learning in BNs is that the learning datasets are seldom complete and one have to deal with missing observations. Inference with missing data is an old problem in statistics and several solutions have been proposed in the last three decades starting from the pioneering work of (Dempster et al., 1977). These authors proposed a famous algorithm that iterates, until convergence towards stationary point, between two steps, one called Expectation or E-step in which the expected values of the missing data are inferred from the current model parameter configuration and the other, called Maximization or M-step, in which we look for and find the parameter values that maximize a probability function (e.g. likelihood). This algorithm, known as the Expectation-Maximization (or EM) algorithm has become a routine technique for parameters estimation in statistical models with missing data in a wide range of applications. Lauritzen, (1995) described how to apply the EM algorithm to learn parameters for known structure BNs using either Maximum-Likelihood (ML) or maximum a posteriori (MAP) estimates (so called EM-MAP) (McLachlan et al., 1997).

Learning structure (graphical structure of conditional dependencies) in BNs is a much more complicated problem that can be formally presented in classical statistics as a model selection problem. In fact, it was shown that learning structure from data is an NP-hard problem

(Chickering et al., 2004) and that the number of structures for a given number of nodes is super-exponential (Robinson, 1977), making the exploration of the space of all possible structures practically infeasible. Structure learning in BNs has been the subject of active research in the last five years, boosted by the application to high-throughput data in biology, and different heuristics have been proposed. Two major classes of methods can be distinguished; those based on optimizing a score function (finding the structure that maximizes the joint probabilities of the network or some function of it) and those based on correlations (see Leray, (2006) for a review).

Hassairi et al., (2005) have proposed a new inference framework in statistical models that they named "Implicit inference". Implicit inference can be shortly defined as "Bayesian inference without priors" which seems like a nonsense at first sight. In fact, Implicit inference derives a special kind of posterior distribution (called Implicit distribution) that corresponds to an improper choice of the prior distribution (see details below). We recently applied this new Implicit inference framework to learning parameters in BNs with complete (Ben Hassen et al., 2008) and incomplete data (Ben Hassen et al., 2009). In this last work, a novel algorithm, similar to EM (that was called I-EM) was proposed and was shown to have better convergence properties compared to it. For structure learning in BNs, we also proposed a new score function (Implicit score) and implemented it within well known algorithms (Bouchaala et al., 2010).

In this chapter, we give a thorough presentation of the Implicit method applied to parameters and structure learning in BNs and discuss its advantages and caveats. An example application is given to illustrate the use of our method.

2. Inference with the Implicit Method

2.1 A quick tour in the Implicit world

The basic idea of the Bayesian theory is to consider any unknown parameter θ as a random variable and to determine its posterior (conditional) distribution given data and an assumed prior distribution (see for example Robert, 1994). The choice of a prior is generally based on the preliminary knowledge of the problem.

Recently, Hassairi et al., (2005) introduced the concept of Implicit distribution which can be described as a kind of posterior distribution of a parameter given data. To explain the principle of Implicit distribution let us consider a family of probability distributions $\{p(x/\theta), \theta \in \Theta\}$ parameterized by an unknown parameter θ in a set Θ, where x is the observed data.

The Implicit distribution $p(\theta/x)$ is calculated by multiplying the likelihood function $p(x/\theta)$ by a counting measure σ if Θ is a countable set and by a Lebesgue measure σ if Θ is an open set (σ depends only on the topological structure of Θ) and then dividing by a norming constant $c(x) = \int_{\Theta} p(x/\theta)\sigma(d\theta)$. Therefore the Implicit distribution is given by the following formula $p(\theta/x) = (c(x))^{-1}p(x/\theta)\sigma(\theta)$ and plays the role of a posterior distribution of θ given x in the Bayesian method, corresponding to a particular improper prior which depends only on the topology of Θ (without any statistical assumption). The Implicit distribution, which exists for most (but not all) statistical models, can be used for the estimation of the parameter θ following a Bayesian methodology. In fact, the Implicit estimator $\widehat{\theta}$ of θ corresponds to the mean (first moment) of the Implicit distribution.

2.2 A simple example: Implicit estimation in binomial distribution case

To illustrate how the Implicit method proceeds let us consider a simple example. Let $X = (N_1, N_2)$ be a random variable following a binomial distribution with unknown parameters $N = N_1 + N_2$ and $\theta = (\theta_1, \theta_2)$. We first estimate N by the Implicit method after that we use the estimate $\widehat{N}$ to estimate θ. After some calculations, we obtain

$$P(N/X) = \frac{P(X/N)}{C(X)} = C_N^{\overset{\vee}{N_1}} \theta_1^{N - \overset{\vee}{N_1}} (1 - \theta_1)^{\overset{\vee}{N_1} + 1},$$

where $\overset{\vee}{N_1} = N - N_1 = \sum_{i=2}^{r} N_i$.

So, the Implicit distribution of N given $X = (N_1, ..., N_r)$ is a Pascal distribution with parameters $1 - \theta_1$ and $\overset{\vee}{N_1} + 1$. Suppose that θ_1 is known, the Implicit estimator $\widehat{N}$ of N is the mean of the Pascal distribution:

$$\widehat{N} = E(N/X) = \sum_{N \geq 0} N C_N^{\overset{\vee}{N_1}} \theta_1^{N - \overset{\vee}{N_1}} (1 - \theta_1)^{\overset{\vee}{N_1} + 1}.$$

Let N_{ob} be the number of observations and take

$$\theta_{k_0} = \max\{ \frac{N_k}{N_{ob}}; \ \frac{N_k}{N_{ob}} \leq \frac{1}{r-1} \text{ and } 1 \leq k \leq r \}.$$

After some calculations, we have

$$\widehat{N} = \frac{(\overset{\vee}{N_{k_0}} + 1)}{1 - \theta_{k_0}} = N_{ob} + \frac{N_{k_0}}{\overset{\vee}{N_{k_0}}},$$

where $\overset{\vee}{N_{k_0}} = N_{ob} - N_{k_0}$

Consequently, the probability of the next observation to be in state x^k given a dataset D is obtained by

$$\widehat{\theta}_k = P(X_{N_{ob}+1} = x^k / D) = \frac{N_k + 1}{\widehat{N} + r}, 1 \leq k \leq r \text{ and } k \neq k_0 \tag{2.1}$$

and $\widehat{\theta}_{k_0} = 1 - \sum_{i \neq k_0} \widehat{\theta}_i$

other examples and selected applications of Implicit distributions can be found in the original paper (Hassairi et al., 2005).

2.3 Implicit inference with Bayesian Networks

Formally, a Bayesian network is defined as a set of variables $X = \{X_1, ..., X_n\}$ with :(1) a network structure S that encodes a set of conditional dependencies between variables in X, and (2) a set P of local probability distributions associated with each variable. Together, these components define the joint probability distribution of X.

The network structure S is a directed acyclic graph (DAG). The nodes in S correspond to the variables in X_i. Each X_i denotes both the variable and its corresponding node, and $Pa(X_i)$ the parents of node X_i in S as well as the variables corresponding to those parents. The lack of

possible arcs in S encode conditional independencies. In particular, given structure S, the joint probability distribution for X is given by the product of all specified conditional probabilities:

$$P(X_1, ..., X_n) = \prod_{i=1}^{n} P(X_i / Pa(X_i)) \tag{3.1}$$

a factorization that is known as *the local Markov property* and states that each node is independent of its non descendant given the parent nodes. For a given *BN* the probabilities will thus depend only on the structure of the parameters set.

3. Learning parameters from complete data

In this section we consider the learning of parameters in BNs with discrete variable, that is for every node i the associated random variable X_i takes r_i states :

$$node\ 1 \rightarrow X_1 \in \{x_1^1, ..., x_1^{r_1}\}$$

$$node\ 2 \rightarrow X_2 \in \{x_2^1, ..., x_2^{r_2}\}$$

$$\vdots$$

$$node\ i \rightarrow X_i \in \{x_i^1, ..., x_i^{r_i}\}$$

$$\vdots$$

$$node\ n \rightarrow X_n \in \{x_n^1, ..., x_n^{r_n}\}.$$

Let D be a dataset and let N_{ijk} be a number of observations in D for which the node i is in state k and its parents are in state j that is $X_i = x_i^k$ and $Pa(X_i) = x_i^j$. Note that, since each node might have two or more parents, state j corresponds to a combination of states of the parents. For example if a node has three parents, each having three states, then there are 27 states of the parents and j takes values from 1 to 27.

The distribution of X_i is multinomial with parameters N_{ij} and $\theta_{ij} = (\theta_{ij2}, ..., \theta_{ijr_i})$, where $N_{ij} = \sum_{k=1}^{r_i} N_{ijk}$ and $\theta_{ijk} = P(X_i = x_i^k / Pa(X_i) = x^j)$; $k = 1, ..., r_i$ and $\sum_{k=1}^{r_i} \theta_{ijk} = 1$

$$P(X_i = (N_{ij1}, ..., N_{ijr_i}) / Pa(X_i) = x^j) = N_{ij}! \prod_{k=1}^{r_i} \frac{\theta_{ijk}^{N_{ijk}}}{N_{ijk}!}.$$

Then N_{ij} and θ_{ij} are unknown parameters that will be estimated by the Implicit method. Given a network S, consider for node i, N_{ijob} is the observed number of occurrences of the node i and its parents are in the state j.

Let $\theta_{ijk(0)} = \frac{N_{ijk(0)}}{N_{ijob}} = \max\{\frac{N_{ijk}}{N_{ijob}}; \quad \frac{N_{ijk}}{N_{ijob}} \leq \frac{1}{r_i - 1} \text{ and } 1 \leq k \leq r_i\}$.

The application of the Implicit method gives the following estimation of N_{ij} and θ_{ij}:

$$\widehat{N}_{ij} = N_{ijob} + \frac{N_{ijk(0)}}{\overset{\vee}{N}_{ijk(0)}}; \tag{3.2}$$

where $\overset{\vee}{N}_{ijk(0)} = N_{ijob} - N_{ijk(0)}$ and

$$\widehat{\theta}_{ijk} = \frac{N_{ijk} + 1}{\widehat{N}_{ij} + r_i} \ \text{ if } \ k \neq k(0) \tag{3.3}$$

and

$$\widehat{\theta}_{ijk(0)} = 1 - \sum_{k \neq k(0)} \widehat{\theta}_{ijk}$$

4. Learning parameters from incomplete data

Consider a dataset D with missing data, we compute the Implicit distribution $P(\theta/D)$ and use the distributions in turn to compute expectation of parameters of interest. Let X be a random variable that follows a multinomial distribution with parameters N and $\theta = (\theta_1, ..., \theta_r)$ such that $Y = (N_1, ..., N_r) \subset X$ and $Z = (N_1^*, ..., N_r^*) \subset X$ denote the observed and unobserved variables, respectively. So, $X = (N_1 + N_1^*, ..., N_r + N_r^*)$
and $P(\theta/Y) = \sum_Z P(Z/Y)P(\theta/Y, Z)$

To estimate the parameters θ_{ijk} of the network, with incomplete dataset, we propose a new iterative algorithm named Implicit EM (or in short I-EM) algorithm. Consider a node i with parents in the state j and a dataset D which contains $N_{ij}^{(0)}$ observed and unobserved values in such state. Let $N_{ijob}^{(0)}$ the observed values in D, so $N_{ij}^{(0)} > N_{ijob}^{(0)}$ and $N_{ij}^{(0)} - N_{ijob}^{(0)}$ represents the number of unobserved states.

So, the initial conditions for a node i are:

$N_{ij}^{(0)}$ is the number of observed and unobserved states.

$\theta_{ijk}^{(0)}$ is the observed frequency of the node i in the state k given its parents in the state j. Then, $N_{ijk}^{(0)} = N_{ij}^{(0)} \theta_{ijk}^{(0)}$ is the number of observed occurrences of the node i in the state k and its parents in the state j.

$$N_{ijob}^{(0)} = \sum_{k=1}^{r_i} N_{ijk}^{(0)}$$

The I-EM algorithm is iterative and involves three steps; the first step consists in getting the maximum of the conditional frequencies, the second step estimates the number of observations from the first step and the third computes the other conditional probabilities. Formally, the algorithm iterates through the following steps, until convergence:

(1) Choose the maximum frequency $k(0)$

(2) Estimate the number of observations $N_{ij}^{(1)}$

(3) Compute the conditional probabilities $\theta_{ijk}^{(1)}$

with the
stop condition being:

Compute the sum of estimated occurrences $\sum_{k=1}^{r_i} N_{ijk}^{(t)}$

if $\sum_{k=1}^{r_i} N_{ijk}^{(t)} > N_{ij}^{(0)}$ then stop, otherwise continue steps (1) to (3).

The philosophy of our algorithm is to virtually fill the missing data for all nodes until all missing cells in the database are completed. A detailed description and a formal proof of convergence of the I-EM algorithm is given in (Ben Hassen et al., 2009).

5. Learning Bayesian Network Structure

Learning Bayesian Network structure from database is an NP-hard problem and several algorithms have been developed to obtain a sub-optimal structure from a database. Most of the widely used methods are score metric-based methods. By these methods a scoring metric is defined and computed for each candidate structure and a search strategy (algorithm) is used to explore the space of possible, alternative structures and identify the one (or those) having the highest score.

5.1 Score metrics

A scoring criteria for a DAG is a function that assigns a value to each DAG based on the data. Cooper and Hersovits (1992) proposed a score based on a Bayesian approach with Dirichlet priors(known as BD: Bayesian Dirichlet). Starting from a prior distribution on the possible structure $P(B)$, the objective is to express the posterior probability of all possible structures ($P(B|D)$ or simply $P(B, D)$) conditional on a dataset D:

$$S_{BD}(B, D) = P(B, D) = \int_{\Theta} P(D|\Theta, B) P(\Theta|B) P(B) d\Theta = P(B) \int_{\Theta} P(D|\Theta, B) P(\Theta|B) d\Theta$$

The BD score is analitycally expressed as:

$$S_{BD}(B, D) = P(B) \prod_{i=1}^{n} \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(N_{ij} + r_i - 1)!} \prod_{k=1}^{r_i} N_{ijk}! \tag{5.2}$$

The BIC (Bayesian Information Criteria) score metric was proposed by Schwartz (1978) and is defined as:

$$S_{BIC} = log L(D|\theta^{MV}, B) - \frac{1}{2} Dim(B) log N \tag{5.3}$$

where θ^{MV} is the maximum likelihood estimate of the parameters, B is the BN structure and $Dim(B)$ is the dimension of the network defined by : $Dim(B) = \sum_{i=1}^{n} Dim(X_i, B)$ and $Dim(B) = (r_i - 1)q_i$

Another common score in structure learning is the Mutual Information (MI). The Mutual Information between two random variables X and Y, denoted by $I(X, Y)$ is defined by Chow and Liu (1968):

$$I(X, Y) = H(X) - H(X|Y) \tag{5.4}$$

Where $H(X)$ is the entropy of random variables X defined as:
$H(X) = - \sum_{i=1}^{r_x} P(X = x_i) log(P(X = x_i))$
and
$H(X|Y) = - \sum_{i=1}^{r_x} \sum_{j=1}^{r_y} P(X = x_i / Y = y_j) log(P(X = x_i | Y = y_j))$ where r_x and r_y are the number of discrete states for variables X and Y, respectively.

5.2 Algorithms for structure learning

One of the most used algorithms is the K2 algorithm (Cooper and Herskovits (1992). This algorithm proceeds as follows: we assume an initial ordering of the nodes to reduce computational complexity and assume that the potential parent set of node X_i can include only those nodes that precede it in the input ordering.

Chow et al., (1968) proposed a method derived from the Maximum Weight Spaning Tree (MWST). This method associates a weight to each potential edges $X_i - X_j$ of the tree. This weight may be the MI(equation 5.4), or the local variation of the score proposed by (Heckerman et al., 1994). Given the weight matrix, we can use the Kruskal algorithm (Kruskal 1956) to obtain a directed tree by choosing a root and then browsing the tree by an in-depth search. The GS (Greedy Search) algorithm takes an initial graph, then associates a score for each neighborhood. The graph with the highest score in this neighborhood is then chosen as the starting graph for the next iteration.

5.3 The Implicit Score (IS)

The Implicit Score(IS) have the same derivation as the the BD score in which the Implicit estimators of the paremeters (see equations 3.2 and 3.3) are used rather than Bayesian estimators (Bouchaala et al., 2010). The expression of the Implicit score (IS) is thus obtained by substituting in equation 5.2 N_{ijk} by $\widehat{N}_{ijk}\widehat{\theta}_{ijk}$ and N_{ij} by $\widehat{N}_{ij}$:

$$S_{IS}(B, D) = P(B) \prod_{i=1}^{n} \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(\widehat{N}_{ij} + r_i - 1)!} \prod_{k=1}^{r_i} \widehat{N}_{ij}\widehat{\theta}_{ijk}! \tag{5.5}$$

.

We implemented this score within K2, MWST and GS algorithms for network structure learning. Performance of IS was evaluated on a benchmark database (ASIA network (lauritzen and Spiegelhater, 1988) in comparison to other score metrics, namely BIC, BD and MI.

The experiments were carried out on different datasets randomly selected from the ASIA database (20,000 data points). The dataset size was varied from 100 to 1000 (in order to test robusteness to small databases)and 20 replicates were performed for each database size. The performance of each score was evaluated by four criteria : the average (over the replicates) numbers of missings edges, additional edges, reversed edges and correct edges (relative to the true structure inferred from the whole database).

Table 1 below shows that the Implicit score yields improved performance over other scores when used with the MWST and GS algorithm, and have similar performance when implemented within K2 algorithm.

6. Application to real data: thyroid cancer prognosis

To illustrate how the Implicit method proceed, we consider an example on thyroid cancer. The dataset comprises data on 92 thyroid cancer patients described in Rebai et al., (2009a,b). We considered only five nodes with two states each:

Therapeutic response (TR): no response (1)/complete remission (2)

Metastasis (MET) yes (1)/no (2).

Thyroglobulin level (TG) low: ≤ 30 ng/mL (1); high: > 30 ng/mL (2)).

The genotype of a single nucleotide polymorphism within the HER2 gene (HER2): genotype AA(1); genotype AG (2)(here genotype GG was totally absent).

The genotype of a single nucleotide polymorphism within the estrogen receptor gene (ER): genotype AA and AG(1); genotype GG (2) (note here that genotypes AA and AG were merged

MWST Algorithm	IS	BIC	MI	Best Result	
Correct Arc	4,39	2,62	2,71	8	
Reversed Arc	1,93	3,08	3,08	0	(A)
Missing Arc	1,68	2,3	2,21	0	
Extra Arc	0,68	1,32	1,22	0	

K2 Algorithm	IS	BIC	BD	Best Result	
Correct Arc	4,66	4,7	4,88	8	
Reversed Arc	1,59	1,69	1,71	0	(B)
Missing Arc	1,75	1,61	1,41	0	
Extra Arc	1,51	1,34	1,85	0	

GS Algorithm	BIC-BIC	MI-BD	IS-BIC	IS-BD	Best Result	
Correct Arc	4,18	4,08	5,28	5,42	8	
Reversed Arc	1,92	2,34	0,82	0,92	0	(C)
Missing Arc	1,9	1,58	1,9	1,66	0	
Extra Arc	0,88	1,82	0,62	1,26	0	

Table 1. Comparative Analysis of the Implicit score (IS) with BD, BIC and MI scores implemented within (A) MWST algorithm, (B)K2 algorithm and (C) GS algorithm.

together because A is a risk allele). These two polymorphisms were included due to their highly significant association, inferred by bivariate and multivariate statistical tests, with the three other variables (see Rebai et al., 2009b for more details on the data).

The structure obtained by the K2 algorithm with the Implicit score is given in figure 1. Note that the same structure was obtained by the BD score.
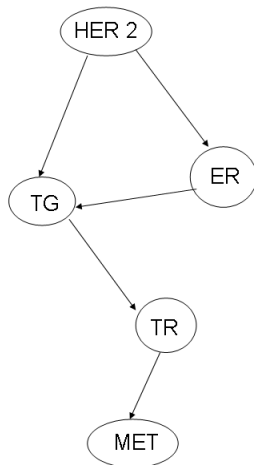

Fig. 1. The structure obtained by the K2 algorithm with the Implicit score

Using this structure we estimated the parameters by the Implicit approach. For parameter notations, nodes are denoted as: (1)ER, (2)HER2, (3)TG, (4)TR and (5)MET. Parameter t_{ijk} corresponds to the node i in state k and its parents in state j. According to the structure in figure1, one node (HER2) has no parents, three nodes have one parent and one node has two

parents (TG). Consequently we have two parameters for HER2, four for ER, TR and MET and eight for TG.

parameter	estimated value	parameter	estimated value
t111	0.20608440	t112	0.79391560
t121	0.50137741	t122	0.49862259
t211	0.72054405	t212	0.27945595
t311	0.47058824	t312	0.52941176
t321	0.53736875	t322	0.46263125
t331	0.43298969	t332	0.56701031
t341	0.85161290	t342	0.14838710
t411	0.21479714	t412	0.78520286
t421	0.94267026	t422	0.05732974
t511	0.07434944	t512	0.92565056
t521	0.92560895	t522	0.07439105

Table 2. Parameters Estimates from a complete dataset of 94 thyroid cancer patients based on structure in Fig1.

If we look at the TR node and particularly the probability of the occurrence of a positive response to therapy (t412) we see that it is high (almost 80 %) when the parent (TG) is at state 1, that is for patients with low TG levels while it is small (about 6 %) for patients with high TG levels (t422). This confirm the high prognostic value of TG level, well recognized by clinicians. Another expected result is that the probability of having metastasis is very high (92 %) when the patient does not respond to therapy (t512). However, an original result is that the probability of having a high TG levels is small (about 15 %) when the patient carries non-risk genotypes at the two single nucleotide polymoprhisms (t342)compared to corresponding probabilities to carriers of a risk genotype for at least one SNP (50 % on average). This means that the two SNP can be used as early prognostic factors that predict the increase in TG levels, which might be of help for therapeutic adjustment (preventive treatment,..).

In order to test the robustness of the Implicit method in parameter learning, we introduced 5 % missing data by randomly deleting 5 % of the data for each node. Table 3 gives the parameters estimates and shows that the change in parameters estimates is slight except for the node without parents (HER2). This property of Implicit estimators has already been reported in Ben Hassen et al., (2009) and is expected because nodes without parents are expected to be more sensitive to missing data.

7. Conclusion

In this chapter, we described the Implicit method, a new framework for learning structure and probabilities in Bayesian networks. We showed how our method proceeds with complete and incomplete data. The use of the Implicit method was illustrated on a real and original dataset of thyroid cancer.

The Implicit method is a new approach that can be seen as a prior-free Bayesian approach. It has the advantages of Bayesian methods without their drawbacks. In fact, the choice of prior information in Bayesian approaches has always been problematic and has been advanced by many critics to be the major weakness of such methods. Implicit method avoids the problem of priors and leads to estimators and algorithms that are easier to derive and to implement.

We showed here and in our previous work that the Implicit score when implemented within

parameter	estimated value	parameter	estimated value
t111	0.2173913	t112	0.7826087
t121	0.3333333	t122	0.6666667
t211	0.7437071	t212	0.2562929
t311	0.4615385	t312	0.5384615
t321	0.5381062	t322	0.4618938
t331	0.4545455	t332	0.5454545
t341	0.9166667	t342	0.08333333
t411	0.2005571	t412	0.7994429
t421	0.9589041	t422	0.04109589
t511	0.0787401	t512	0.9212598
t521	0.948718	t522	0.05128205

Table 3. Table of estimated parameters for a 5 % rate of missing data for thyroid cancer patients

traditional algorithms for structure learning (and particularly the MWST algorithm) leads to better results and seems to be more robust when the database is of relatively small size. This might be a very useful property for applications in medical prognosis or diagnosis of rare diseases, where the number of patients has been a limiting factor to the use of Bayseian networks for modeling the complex relationship between several predicting factors, such as clinical, molecular, biochemical and genetical factors.

The easy implementation of the Implicit algorithm for parameters learning in Bayseian networks with missing data and its performance compared to the EM algorithm and particularly its faster convergence, is one of the reasons that can lead to its adoption for many applications in computational biology and genomics (see Needham et al., 2007).

In its current version, the Implicit method can only handle Bayesian networks with discrete variables. This of course encloses a wide range of applications, but the generalization to networks with continuous or mixed variables is our next challenge and will be addressed in the near future.

Acknowledgments

8. References

Ben Hassen, H., Masmoudi, A. and Rebai, A., 2008. Causal inference in Biomolecular Pathways using a Bayesian network approach and an Implicit method. *J. Theoret. Biol.* 4, 717-724.

Ben Hassen, H., Masmoudi, A. and Rebai, A., 2009. Inference in signal transduction pathways using EM algorithm and an Implicit Algorithm : Incomplete data case. *J. comp. Biol.* 16, 1227-1240.

Bouchaala, L., Masmoudi, A., Gargouri, F and Rebai, A., 2010. Improving algorithms for structure learning in Bayesian Networks using a new Implicit score. *Expert. Syst. Appl..* 37, 5470-5475.

Chickering, D., Heckerman, D. and Meek, C., 2004. Large-Sample Learning of Bayesian Networks is NP-Hard. *J. Mach. Learn. Res.* 5, 1287-1330.

Cooper, G.F. and Herskovits, E., 1992. A Bayesian method for the induction of probabilistic networks from data. *Mach. Learn*. 9, 309-347.

Chow , C. K., Liu, C. N., 1968. Approximating discrete probability distributions with dependence trees. *IEEE T. Knowl. Data. En.*.20, 1-13.

Chrisman, L., 1996. A road map to research on Bayesian networks and other decomposable probabilistic models, *Technical Report, School of Computer Science, CMU, Pittsburgh, PA*.

Cooper, G., Hersovits,E. 1992. Abayesian method for the induction of probabilistic networks from data, *Mach Learn*. 9, 309-347.

Dempster, A.P., Laird, N.M. and Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. Roy. Statist. Soc. B*. 39, 1-38.

Hassairi, A., Masmoudi, A. and Kokonendji, C., 2005. Implicit distributions and estimation. *Comm. Statist. Theory Methods*. 34, 245-252.

Heckerman, D., Geiger, D. and Chickering, M., 1994. Learning Bayesian Networks : The combination of knowledge and statistical data, *Proceedings of the 10th conference on uncertainty in artificial intelligence*. San francisco, CA, USA : Morgan Kaufmann Publishers. 293-301.

Krause, P., 1996. Learning probabilistic networks. *Technical Report, Philips Research Laboratories, UK*.

Kruskal, J., 1956. On the shortest spaning subtree of a graph and traveling salesman problem. *Proceedings of the American Mathematical Society*. 7, 48-50.

Leray, P., 2006. Habilitation a diriger les recherches: *Automatique et Traitement du Signal Réseaux bayésiens : apprentissage et modélisation de systèmes complexes.* Université de Rouen, UFR des Sciences.

Linggi, B. and Carpenter G., 2006. ErbB receptors: new insights on mechanisms and biology. *Trends. Cell. Biol*. 16, 649-656.

Lauritzen, S.L., 1995. "The EM algorithm for graphical association models with missing data." *Comput. Statist. Data Anal*. 19, 191-201.

Lauritzen, S. and Spiegelhater, D., 1988. Local computation with probabilities on graphical structures and their application on expert systems. J. Roy. Stat. Soc. 50, 157-224.

Mamitsuka, H., 2005. Essential Latent Knowledge for Protein-Protein Interactions: Analysis by Unsupervised Learning Approach *IEEE/ACM Trans. Comput. Biol. Bioinform*. 2, 119-130.

McLachlan, G.J. and Krishnan, T., 1997. *The EM algorithm and extensions*. Wiley, New York.

Mukhopadhyay, N., 2006. Some Comments on Hassairi et al.'s "Implicit Distributions and Estimation". *Comm. Statist. Theory Methods*. 35, 293-297.

Nakayama, T., Asai, S., Sato, N. and Soma, M., 2006. Genotype and haplotype association study of the STRK1 region on 5q12 among Japanese: a case-control study. *Stroke*. 37, 69-76.

Needham, C.J., Bradford, J.R., Bulpitt, A.J., Westhead, D.R., 2007. A Primer on Learning in Bayesian Networks for Computational Biology. *PLoS Comput. Biol*. 3, 8, e129.

Normanno, N., De Luca, A., Bianco, C., Strizzi, L., Mancino, M., Maiello, M.R., Carotenuto, A., De Feo, G., Caponigro, F. and Salomon, D.S., 2006. Epidermal growth factor receptor (EGFR) signaling in cancer. *Gene*. 366, 2-16.

Rebai M, Kallel I, Hamza F, Charfeddine S, Kaffel R, Guermazi F and Rebai A. 2009a. Association of EGFR and HER2 Polymorphisms with Risk and Clinical Features of Thyroid Cancer. *Genet. Test. Mol. Biomarkers*, 13: 779-784.

Rebai M., Kallel I, Charfeddine S., Hamza F, Guermazi F., Rebaï A. 2009b. Association of poly-
        morphisms in oestrogen and thyroid hormone receptors with thyroid cancer risk. *J
        Rec. Signal Transd.*, 29: 113-118.
Robinson, R.W,1977. Counting unlabeled acyclic digraphs. *Comb. Mathem.*, 622, 28-43
Robert, C.P., 1994. *The Bayesian Choice: A decision-Theoretic Motivation*. Springer-Verlag, New-
        York.
Schwarz G., 1978. Estimating the dimension of a model. *Ann. Statist.* 6, 461-464.
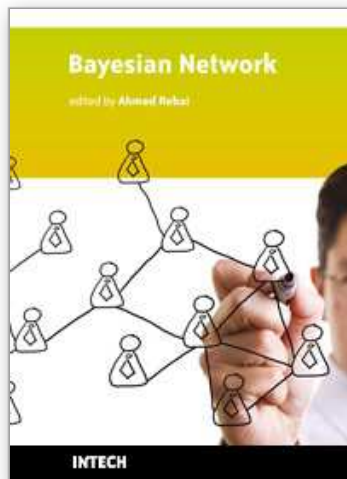
Bayesian Network

Edited by Ahmed Rebai

Bayesian networks are a very general and powerful tool that can be used for a large number of problems involving uncertainty: reasoning, learning, planning and perception. They provide a language that supports efficient algorithms for the automatic construction of expert systems in several different contexts. The range of applications of Bayesian networks currently extends over almost all fields including engineering, biology and medicine, information and communication technologies and finance. This book is a collection of original contributions to the methodology and applications of Bayesian networks. It contains recent developments in the field and illustrates, on a sample of applications, the power of Bayesian networks in dealing the modeling of complex systems. Readers that are not familiar with this tool, but have some technical background, will find in this book all necessary theoretical and practical information on how to use and implement Bayesian networks in their own work. There is no doubt that this book constitutes a valuable resource for engineers, researchers, students and all those who are interested in discovering and experiencing the potential of this major tool of the century.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Hanen BenHassen, Lobna Bouchaala, Afif Masmoudi and Ahmed Rebai (2010). Learning Structure and Parameters in Bayesian Networks Using an Implicit Framework, Bayesian Network, Ahmed Rebai (Ed.), ISBN: 978-953-307-124-4, InTech, Available from: http://www.intechopen.com/books/bayesian-network/learning-structure-and-parameters-in-bayesian-networks-using-an-implicit-framework

INTECH
open science | open minds