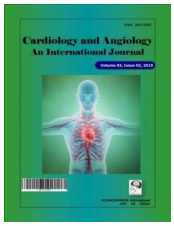

Cardiology and Angiology: An International Journal
4(2): 43-50, 2015, Article no.CA.2015.026
ISSN: 2347-520X

SCIENCEDOMAIN *international*
www.sciencedomain.org

Learning Bayesian Networks Using Heart Failure Data

Muzeyin Ahmed Berarti[1] and Ayele Taye Goshu[2*]

[1]Department of Statistics, Wolaita Sodo University, P.O.Box 138, Ethiopia.
[2]Department of Statistics, School of Mathematical and Statistical Sciences, Hawassa University, P.O.Box 05, Ethiopia.

Authors' contributions

This work was carried out in collaboration between both authors. Author MAB conceptualized, proposed and designed the research, design R-code and existing library routines and software packages in combination, analyzed the data and wrote all drafts of the manuscript. Author ATG commented with dedication on the drafts, participated in editing design of the study, performed the statistical analysis, managed the literature searches and approved the final manuscript. Both authors conventionally read and approved the final manuscript.

Article Information

DOI: 10.9734/CA/2015/17792
Editor(s):
(1) Francesco Pelliccia, Department of Heart, Great Vessels University La Sapienza, Rome, Italy.
Reviewers:
(1) Loc Nguyen, University of Science, Ho Chi Minh, Vietnam.
(2) Cliff Richard Kikawa, Mathematics and Statistics, Tshwane University of Technology, Republic of South Africa.
Complete Peer review History: http://sciencedomain.org/review-history/10072

Original Research Article

Received 26th March 2015
Accepted 19th May 2015
Published 7th July 2015

ABSTRACT

Background: Several factors may affect heart failure status of patients. It is important to investigate whether or not the effects are direct. The purpose of this study was learning Bayesian networks that encode the joint probability distribution for a set of random variables.
Methods: The design was a retrospective cohort study. The target population for this study was heart failure patients who were under follow- up at Asella referral teaching Hospital from February, 2009 to March, 2012. Bayesian Network is used in this paper to examine causal relationships between variables via Directed Acyclic Graph (DAG).
Results: Death of patients can be determined using HIV, hypertension, diabetes, anemia, renal inefficiency and sinus. Hypertension and sinus were found to have direct effects while TB had only indirect effect. Age did not have an effect.
Conclusion: Anemia, HIV, diabetes mellitus renal inefficiency and sinus directly affect the death of heart failure patient. Death is conditionally independent on TB and age, given all other variables.

Keywords: Bayesian network; parameter learning; structure learning; causal relationships.

**Corresponding author: E-mail: ayele_taye@yahoo.com;*

1. INTRODUCTION

Graphs, as stated in [1], provide a comprehensive picture of a problem that makes for a more complete and better balanced understanding than could be derived from tabular or textual forms of presentation. Moreover, graphs can bring out hidden facts and relationships and can stimulate, as well as aid, analytical thinking and investigation.

The main focus of this paper is to describe and drive conditional independence relations existing among random variables through Directed Graphical model called Bayesian networks. Bayesian networks are among the leading techno logies to investigate such relations.

A Bayesian network [2] is a graphical model that encodes the joint probability distribution for a set of random variables and a way of finding important relationships between variables. Bayesian Networks provide a power full technique to understand causal relationships between variables via Directed Acyclic Graph (DAG). Directed graphical models represent probability distributions that can be factored into products of conditional distributions, and have the potential for causal interpretations. The nodes in the graph represent the random variables and missing arrows between the nodes, specify properties of conditional independence between the variables. We refer the reader [3] for detailed understanding of Undirected Graphical Models counterpart and comprehensive comparison with Directed Graphical Models.

A method for learning the parameters and structure of such Bayesian networks has recently been described by in [4]. [5] described a modern method for learning the parameters and structures of Bayesian networks in deal package of R statistical Software.

Bayesian networks are designed for making decisions in systems with uncertainties [6].

Bayesian networks are therefore suitable for problems where the variables exhibit a complicated dependency structure.

2. METHODS

2.1 Data Description

The design was a retrospective cohort study, which reviews the patient's card and information sheet. The data of size 263 were obtained from record reviews of all inpatient heart failure patients admitted to Asella Referral Hospital from February, 2009 to March, 2012.

2.2 Response Variables

Death status of patients during hospital stays due to heart failure. This status of patient is coded as 1 if the patient died in hospital and 0 if the patient alive.

2.3 Independent Variables

The prognostic variables which are expected to be the risk factors of heart failure are categorical and continuous (see Table 1).

Table 1. Independent variables and their coding

Variables	Category and coding
Age at the start of treatment	Continues
Renal inefficiency	Yes= 1, No= 0
TB	Positive= 1, Negative= 0
Diabetes mellitus	Positive= 1, Negative= 0
HIV	Reactive= 1, Nonreactive= 0
Anemia	Anemic = 1, Non-anemic = 0
Hypertension	Positive= 1, Negative= 0
Sinus	Positive= 1, Negative= 0

2.4 Bayesian Networks

A Bayesian network is a graphical model that encodes the joint probability distribution for a set of random variables. Bayesian network, [7], is a specific type of graphical model which is a directed acyclic graph (DAG). [8] described A Bayesian network as a compact, graphical model of a probability distribution. It consists of two parts: a directed acyclic graph which represents direct influences among variables, and a set of conditional probability tables that quantify the strengths of these influences. A graph consists of a set of vertices (nodes), along with a set of edges joining some pairs of the vertices. However, the edges have directional arrows (but no directed cycles) [3]. The nodes in the graph represent the random variables and missing arrows between the nodes, specify properties of conditional independence between the variables. Bayesian networks are designed for making decisions in systems with uncertainties. Here we

perform the analysis using Bayesian networks for discrete and continuous variables in which the joint distribution of all the variables are Conditional Gaussian (CG).

Let D = (*V*, *E*) be a Directed Acyclic Graph (DAG), where *V* is a finite nonempty set of nodes and *E* is a finite set of directed edges (arrows) between the nodes. The DAG defines the structure of the Bayesian network. To each node $v \in V$ in the graph corresponds to a random variable X_v. The set of variables associated with the graph *D* is then $X = (X_v)$, $v \in V$. Often, we do not distinguish between a variable X_v and the corresponding node *v*. To each node *v* with parents *pa*(*v*), a local probability distribution, $p(x_v \mid x_{pa(v)})$ is attached. The set of local probability distributions for all variables in the network is *P*. A Bayesian network for a set of random variables *X* is then the pair (*D*, *P*). The possible lack of directed edges in *D* encodes conditional independences between the random variables *X* through the factorization of the joint probability distribution,

$$p(X) = \prod_{v \in V} p(x_v \mid x_{pa(v)}) \quad (1)$$

Here, Bayesian networks with both discrete and continuous variables are allowed as treated in [4] So the set of nodes *V* is given by $V = \Delta \cup \Gamma$, where Δ and Γ are the sets of discrete and continuous nodes, respectively. The corresponding random variables *X* can then be denoted by

$$X = X_{(v)}, v \in V = (I, Y) = ((I_\delta), \delta \in \Delta, (Y_\gamma), \gamma \in \Gamma$$

,i.e. I and *Y* used for the sets of discrete and continuous variables, respectively. The set of levels for each discrete variable $\delta \in \Delta$ is denoted as I_δ. To ensure availability of exact local computation methods, we do not allow discrete variables to have continuous parents. The joint probability distribution then factorizes into a discrete part and a mixed part, so

$$p(x) = p(i, y) = \prod_{\delta \in \Delta} p(i_\delta \mid i_{pa(\delta)}) \prod_{\gamma \in \Gamma} p(y_\gamma \mid y_{pa(\gamma)}, i_{pa(\gamma)})$$

Where $i_{pa(\gamma)}$ and $y_{pa(\gamma)}$ denote observations of the discrete and continuous parents respectively. A method for estimating the parameters and learning the dependency structure of a conditional Gaussian networks with mixed variables is presented in [4] and implemented in the software package deal in [5] and [9].

2.4.1 Inference

A substantial feature of Bayesian networks is that it enables us to infer conditional dependencies between variables by visually inspecting the network's graph. Therefore we can divide the set of Bayesian network nodes into nonoverlapping subsets of conditional independent nodes. Decomposition is very important when doing inference. Inference is the task of computing the probability of each state of a node in a Bayesian network when other variables are known. To perform inference we first need to be familiar with the belief propagation. Belief propagation is the action of updating the beliefs in each variable when observations are given to some of the variables. Inference in Bayesian networks is performed using Bayes' theorem. Variables in BNs can be divided into groups depending on their position in BNs and taking into account the meaning of real world state that they represent including their observability. Consider a network for a set of random variables X and assume that some of the variables, B, are observed (visible variable) and the rest, A, are not (hidden variable). Let *Uk* be any arbitrary subset of X. The goal of inference is to find the conditional probability density functions (pdfs) over U given the observed variable B Which can be written using Bayes' theorem as

$$p(U_k \mid B) = \frac{p(U_k, B)}{p(B)} = \frac{p(U_k)\, p(B \mid U_k)}{p(B)}$$

Thus $p(U_k)$ is the prior distribution of U_k , i.e. the distribution of U_k before we observe B, $p(B \mid U_k)$ is the likelihood of U_k and $p(B \mid U_k)$ is the posterior distribution of U_k , i.e. the distribution of U_k , when we have observed B.

Generally, finding these distributions are computationally demanding as it involves calculating huge joint distributions, especially if there are many variables in the network. The marginal or conditional distributions of interest can then be found by a series of local computations, involving only some of the

variables at a time. For a thorough treatment of these methods see [2].

2.4.2 Parameter and structure learning

To estimate the parameters in the network and to find the structure of the network, a Bayesian approach has been used. So, regarding the parameters, uncertainty about μ is encoded in a prior distribution $p(\theta)$, using data d to update this distribution (see equation 2), i.e. learn the parameters, and here by obtain the posterior distribution $p(\theta \mid data)$. The section is based on [4] and [10]. Consider a situation with one random variable X. Let θ be the parameter to be assessed and Θ be the parameter space and d a random sample of size n from the probability distribution $p(x \mid \theta)$. We call d our database and $x^c \in d$ a case. Then, according to Bayes' theorem,

$$p(\theta \mid d) = \frac{p(d \mid \theta)\, p(\theta)}{p(d)}, \theta \in \Theta \qquad (2)$$

where $p(d \mid \theta) = \prod_{x^c \in d} p(x^c \mid \theta)$ is the joint probability distribution of *d*, also called the likelihood of *θ*. As prior parameter distributions, the Dirichlet distribution and the Gaussian inverse Gamma distribution have beenused for the discrete variables and for the continuous variables respectively. These distributions are conjugate to observations from the respective distributions and this ensures simple calculations of the posterior distributions. Now, to learn the structure of the network, we calculate the posterior probability o f the D AG, $p(D \mid d)$, which from Bayes' theorem is given by

$$p(D \mid d) = \frac{p(d \mid D)\, p(D)}{p(d)}$$

Where $p(d \mid D)$ is the likelihood of *D* and *p*(*D*) is the prior probability of *D*. As the normalizing constant *p*(*d*) does not depend upon structure, another measure, which gives the relative probability, is

$$p(D, d) = p(d \mid D)\, p(D)$$

We use the above measure and refer to it as the network score. For simplicity, we choose to let $p(D)$ be the same for all DAGs, so we are only interested in calculating the likelihood $p(d \mid D)$. So learning the DAG from data, we can in principle first calculate the network scores for all possible DAGs and then select the DAG with the highest network score. If many DAGs are possible, it is computationally infeasible to calculate the network score for all these DAGs. In this situation it is necessary to use some kind of search strategy to find the DAG with the highest score. In some cases it can be more accurate to average over the possible DAGs for prediction, instead of just selecting a single DAG. So if *x* is the quantity we are interested in, we can use the weighted average,

$$p(x \mid d) = \sum_{D \in DAG} p(x \mid d, D)\, p(D \mid d),$$

Where *D AG* is the set of all DAGs and $p(D \mid d)$ is the weight. Again, if many DAGs are possible, this sum is too hard to compute, so instead, by using a search strategy, we can find a few DAGs with high score and average over these. In order to calculate the network score for a specific DAG D, in a CG network, we need to know the prior probability and the likelihood of the DAG. For simplicity, we could for example choose to let all DAGs be equally likely, then

$$p(D|d) \propto p(d|D)$$

In a CG network, the likelihood of the DAGD is given by

$$p(d \mid D) = \int_{\theta \in \Theta} p(d \mid \theta, D)\, p(\theta \mid D) d\theta,$$

To evaluate which DAG or possible several DAGs that represent the conditional independences in a Bayesian network well, we want to find the DAG or D AGs with the highest network scores. To calculate these scores, we must specify the local probability distributions and the local prior distributions for the parameters for each network under evaluation. We see in the above equation that it, besides the likelihood of the parameters, also involves the prior distribution over the parameters, $p(\theta \mid D)$. This means that we for each possible DAG have to specify a prior distribution for the parameters. In [1] this method is extended to the mixed case. With this method, the parameter priors for all possible networks can be deduced from one joint parameter prior, called master *prior*. To specify

this master prior, we only have to specify a prior Bayesian network, i.e. a prior DAG and a prior probability distribution, together with a measure of how confident we are in the prior network.

2.5 Ethical Considerations

Ethical clearance was obtained from the Hospital.

3. RESULTS AND DISCUSSION

For this study, the data of heart failure patients which was taken from Asella Referral Hospital is used and analyzed using Bayesian Networks. The data of size 263 were obtained from record reviews of all in patient heart failure patients admitted to Medical ward from February, 2009 to March, 2012. In this analysis, nine variables are considered each containing 263 observations as presented in Table 2.

Table 2. Variables used in this analysis

Node index	Variables
1	Age
2	Hypertension
3	HIV
4	Diabetes
5	TB
6	anemia
7	Sinus
8	Renal inefficiency
9	Death

Here we consider Bayesian networks with both discrete and continuous random variables. We use the deal package of R Statistical Software to analyze the data of heart failure patients obtained from Asella Referral Hospital.

The purpose of analyzing data under this section is to find dependency relations between the variables where the main interest lies in finding out which variables influence the death status of heart failure patients.

Age is continuous variable which may cause the death of heart failure patients. However, in deal continuous parents of discrete nodes are not allowed. Thus, describing such a relation is impossible. A remedial measure is handle death as a continuous variable, even though this is clearly not.

3.1 Specification of a Bayesian Network

Here is the R-code for building Bayesian Network for Heart failure Data.

```
> Data<- read.table("mom.dat",header=T) ##
invoke Data from working Directory
> attach(Data)
> Data<-
data.frame(age,hypertension,HIV,diabetes,TB,a
nemia,sinus,renalineffeciency, death)
> Data$death<-as.numeric(Data$death)
> Data$age<-as.numeric(Data$age)
> Data$anemia <-as.factor(Data$anemia)
> Data$diabetes <-as.factor(Data$diabetes)
> Data$hypertension <-
as.factor(Data$hypertension)
> Data$sinus <-as.factor(Data$sinus)
> Data$TB <-as.factor(Data$TB)
> Data$HIV <-as.factor(Data$HIV)
> Data$renalineffeciency <-
as.factor(Data$renalineffeciency)
```

Hereafter, we are in position to specify a prior Bayesian network. We use the empty DAG as the prior DAG since we have no prior knowledge about specific dependency relations and let the probability distribution of the discrete variables be uniform. The assessment of the probability distribution for the continuous variables is based on data.

```
> library(deal) ## call deal package
> Data.nw<-network(Data) ## specify prior
network
> Data.prior <- jointprior(Data.nw) ## create
joint prior distribution
Imaginary sample size: 256
# banlist for age and HIV as none of
#the other variables can influence these
variables.
> from1<-c(2,3,4,5,6,7,8,9)
> to1<-rep(1,8)
> from2<-c(1,2,4,5,6,7,8,9)
> to2<-rep(3,8)
> from3<-rep(9,6)
> to3<-c(2,4,5,6,7,8)
> banlist<-
matrix(c(from1,from2,from3,to1,to2,to3),ncol=2)
> banlist(Data.nw) <- banlist
```

The ban list is a matrix with two columns. Each row contains the directed edge that is not allowed. The final stage is to learn the parameters in the network and initiate the structural learning using *auto search ()* and *heuristic ().*

```
> Data.nw <-
getnetwork(learn(Data.nw,Data,Data.prior))
> Data.search <-
autosearch(Data.nw,Data,Data.prior,trace=TRU
E)
> Data.heuristic <-
heuristic(getnetwork(Data.search),Data,Data.pr
ior,restart=2,

degree=10,trace=TRUE,trylist=gettrylist(Data.se
arch))
```

NB: The banlist forces:

- Death to be a leaf node (death can only receive arcs) (in our dataset death is variable number 9)
- Age and HIV cannot receive arcs (in our dataset age is variable number1 & HIV is variable number 3).

As can be seen from the Bayesian network plot, Fig. 1, death depends directly on Human Immune deficiency Virus (HIV), hypertension, diabetes, anemia, renal inefficiency and sinus. Hence, given these variables, death is conditionally independent on TB and age.

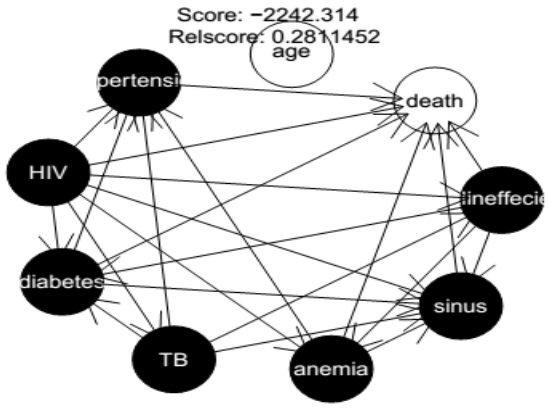

Fig. 1. Bayesian network plot for heart failure data

Death is conditionally independent on TB, given all other variables. We also see that age is independent from all other variables. How can we explain these findings? TB has an impact on death but only indirectly through complications like HIV, hypertension, diabetes, anemia, renal inefficiency and sinus. TB does not provoke death directly but only through these events, which absolutely seems very reasonable. Anemia is one of the important variables that can lead to death directly, but also indirectly: through hypertension and sinus. This shows that anemia has some- how three ways to impact death: directly or indirectly, because it increase the possibility to importance of hypertension and sinus.

As a comparison, we see that HIV has a direct effect and also an indirect effect through all other variables. There are many further interesting indirect effects in this network.

Age seems independent from all other variables. This is an interesting finding of the Bayesian network. Age is clearly associated with death in the other studies (like GAM), but here, when focus is on conditional independence, we estimated that all other variables have an impact on death, direct or indirect, but age has not. We can understand this in the following way: age is in some sense a surrogate variable for health status, in itself age is not a danger, but through events that become more lively with age, like hypertension. Indirect effects are therefore possible, but not direct ones. Apparently there is not enough support in our data to estimate age as a master node that from the top regulates all other variables which then point to death even though age is known as main cause of cardiology.

We build up another separated network that includes age and other causes of heart failure (HIV, hypertension, diabetes, etc) but excludes death in order to discover relationships between age and other causes (HIV, hypertension, diabetes, etc). For brevity, we prefer to determine which variables influence the presence or absence of hypertension. From a medical viewpoint, it is possible that hypertension is among the classical variables related to heart failure and influenced by some of the variables listed in this research.

R-code for new network for presence or absence of hypertension

```
> Data<- read.table("mom.dat",header=T) ##
invoke Data from working Directory
> attach(Data)
> Data2<-
data.frame(age,hypertension,HIV,diabetes,TB,a
nemia,sinus,renalineffeciency)
> Data2$age<-as.numeric(Data$age)
> Data2$anemia <-as.factor(Data2$anemia)
> Data2$diabetes <-as.factor(Data2$diabetes)
> Data$hypertension <-
as.numeric(Data2$hypertension)
```

```
> Data2$sinus <-as.factor(Data2$sinus)
> Data2$TB <-as.factor(Data2$TB)
> Data2$HIV <-as.factor(Data2$HIV)
> Data2$renalineffeciency <-
as.factor(Data2$renalineffeciency)
> library(deal) ## call deal package
> Data.nw2<- network(Data2) ## specify prior
network
> Data.prior2<- jointprior(Data.nw2)
Imaginary sample size: 128
> from1<- c(2,3,4,5,6,7,8)
> to1<- rep(1,7)
> from2<- c(1,2,4,5,6,7,8)
> to2<- rep(3,7)
> from3<- rep(2,5)
> to3<- c(4,5,6,7,8)
> banlist2<- matrix(
c(from1,from2,from3,to1,to2,to3),ncol=2)
> banlist(Data.nw2)<- banlist2
> Data.nw2 <-
getnetwork(learn(Data.nw2,Data2,Data.prior2))
> Data.search2 <-
autosearch(Data.nw2,Data2,Data.prior2,trace=T
RUE)
> Data.heuristic2<-
heuristic(getnetwork(Data.search2),Data2,Data
.prior2,
restart=2,
degree=10,trace=TRUE,trylist=gettrylist(
Data.search2))
```

The network is displayed in Fig. 2. On the contrary to Fig. 1, Fig. 2 implies age is directly related to hypertension which justifies our assumption of indirect impact of age on death status of heart failure patients through hypertension.

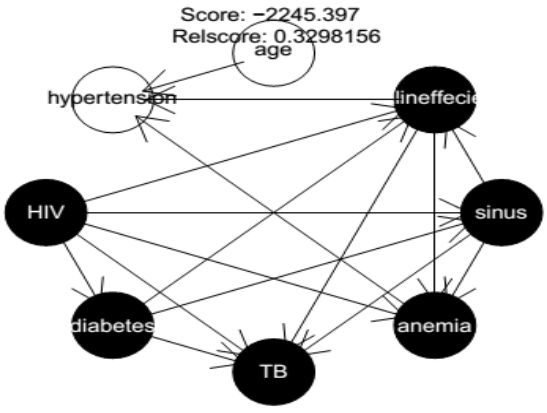

Fig. 2. Network for absence/presence of hypertension

We can build up another separated network that includes age and other causes of heart failure (HIV, hypertension, diabetes, etc) but excludes death status in order to discover dependency relationships between age and other causes (HIV, hypertension, diabetes, etc) in similar manner.

3.2 Discussion

In this paper we have established a nice way of determining the the impact of several variable on death status of heart failure patients. We have given an introduction to Bayesian networks with both discrete and continuous random variables. Several literatures were done on mixed variables (Discrete and continuous). For instance [4,5] and [11]. We also applied Bayesian Networks to clinical data obtained from Asella Referral Hospital. Different literatures have been done on the use of Bayesian Networks in clinical studies one of which is [12].

4. CONCLUSIONS

In this paper we described a powerful technique for analyzing Heart Failure data based on the theory and algorithms for learning Bayesian networks. We explained how to apply these techniques to Heart Failure data. The result of this analysis showed that death of a patient can be determined by HIV, hypertension, diabetes, TB, anemia, renal inefficiency and sinus directly or indirectly. The finding revealed that age does not have direct impact on death of a patient but it has an impact indirectly through complications like hypertension.

CONSENT

It is not applicable.

ACKNOWLEDGEMENTS

The authors would like to express their special gratitude to Professor Arnoldo Frigessi for his useful comments while conducting this study. They also acknowledge Asella Referral Hospital for its cooperation during the time of data collection.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Calvin FS. Handbook of graphic presentation. 2nd ed. A Ronald Press Publication, New York: Wiley. 1954.
2. Cowell RG, Dawid AP, Lauritzen SL, Spiegelhalter DJ. Probabilistic networks and expert systems. Springer-Verlag, Berlin-Heidelberg-New York. 1999.
3. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: Data mining, inference, and prediction. Springer Series in Statistics. 2nd ed. Stanford, California. 2008.
4. Bøttcher SG. Learning bayesian networks with mixed variables. Artificial Intelligence and Statistics 2001, Morgan Kaufmann, San Francisco, CA, USA. 2001;149-156.
5. Bøttcher SG. Learning bayesian networks with mixed variables. Aalborg University. 2003.
   Available: http://www.math.auc.dk/ãlma
6. Koch KR. Introduction to Bayesian statistics. 2nd ed. Springer. 2007.
7. Neapolitan RE. Probabilistic reasoning in expert systems: Theory and algorithms. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York. 1989.
8. Pearl J. Probabilistic reasoning in intelligent systems: Networks of plausible inference. Morgan Kaufmann Publishers, Inc., San Mateo, California. 1988.
9. Bøttcher SG, Dethlefsen C. Learning Bayesian networks with R, In K. Hornik, F. Leisch and A. Zeileis (eds), Proceedings of the 3rd International Workshop on Distributed Statistical Computing. ISSN 1609-395X. 2003b.
10. Heckerman D, Geiger D, Chickering D. Learning Bayesian networks: The combination of knowledge and statistical data. Machine Learning. 1995;20:197-243.
11. Bøttcher SG, Dethlefsen C. Deal: A package for learning Bayesian networks. Journal of Statistical Software. 2003a; 8(20):1-40.
12. Bøttcher SG, Dethlefsen C. Prediction of the insulin sensitivity index using Bayesian network. Department of Mathematical Sciences, Aalborg University Denmark. 2004.

Peer-review history:
The peer review history for this paper can be accessed here:
http://sciencedomain.org/review-history/10072