

Bayesian Networks for enterprise risk assessment

C.E. Bonafede, P. Giudici*

University of Pavia, Italy

Available online 3 March 2007

Abstract

According to different typologies of activity and priority, risks can assume diverse meanings and it can be assessed in different ways.

Risk, in general, is measured in terms of a probability combination of an event (frequency) and its consequence (impact). To estimate the frequency and the impact (severity) historical data or expert opinions (either qualitative or quantitative data) are used. Moreover, qualitative data must be converted in numerical values or bounds to be used in the model.

In the case of enterprise risk assessment the considered risks are, for instance, strategic, operational, legal and of image, which many times are difficult to be quantified. So in most cases only expert data, gathered by scorecard approaches, are available for risk analysis.

The Bayesian Networks (BNs) are a useful tool to integrate different information and in particular to study the risk's joint distribution by using data collected from experts.

In this paper we want to show a possible approach for building a BN in the particular case in which only prior probabilities of node states and marginal correlations between nodes are available, and when the variables have only two states.

© 2007 Elsevier B.V. All rights reserved.

Keywords: Bayesian Networks; Enterprise risk assessment; Mutual information

1. Introduction

A Bayesian Network (BN) is a directed acyclic graph (probabilistic expert system) in which every node represents a random variable with a discrete or continuous state [1,2].

The relationships among variables, pointed out by arcs, are interpreted in terms of conditional probabilities according to Bayes theorem.

With the BN is implemented the concept of conditional independence that allows the factorization of the joint probability, through the Markov property, in a series of local terms that describe the relationships among variables:

$$f(x_1, x_2, \dots, x_n) = \prod_{i=1}^n f(x_i | pa(x_i)),$$

*Corresponding author.

E-mail addresses: bonafede@eco.unipv.it (C.E. Bonafede), giudici@unipv.it (P. Giudici).

URL: <http://www.datamininglab.it> (P. Giudici).

where $pa(x_i)$ denotes the states of the predecessors (parents) of the variable X_i (child) [1–3]. This factorization enable us to study the network locally.

A BN requires an appropriate database (DB) to extract the conditional probabilities (parameter learning problem) and the network structure (structural learning problem) [2,4,5].

The objective is to find the net that best approximates the joint probabilities and the dependencies among variables.

After we have constructed the network one of the common goal of a BN is the probabilistic inference to estimate the state probabilities of nodes given the knowledge of the values of other nodes. The inference can be done from children to parents (this is called diagnosis) or vice versa from parents to children (this is called prediction) [1,4,6].

However, in many cases the data are not available because the examined events can be new, rare, complex or little understood. In such conditions experts' opinions are used for collecting information that will be translated in conditional probability values or in a certain joint or prior distribution (probability elicitation) [5,7–9].

Such problems are more evident in the case where the expert is requested to define too many conditional probabilities due to the number of the variable's parents. So, when possible, it is worthwhile to reduce the number of probabilities to be specified by assuming some relationships that impose bonds on the interactions between parents and children as for example the Noisy-OR and its variation and generalization [2,5,10–12].

In the business field, BNs are a useful tool for a multivariate and integrated analysis of the risks, for their monitoring and for the evaluation of intervention strategies (by decision graph) for their mitigation [2,10,13].

Enterprise risk can be defined as the possibility that something with an impact on the objectives happens, and it is measured in terms of combination of probability of an event (frequency) and of its consequence (impact).

The enterprise risk assessment is a part of Enterprise Risk Management (ERM) where to estimate the frequency and the impact distributions historical data as well as expert opinions are typically used [3,10,13,14]. Then such distributions are combined to get the loss distribution.

In this context BNs are a useful tool to integrate historical data with those coming from experts which can be qualitative or quantitative [9].

2. Our proposal

What we present in this work is the construction of a BN for having an integrated view of the risks involved in the building of an important structure in Italy, where the risk frequencies and impacts were collected by an ERM procedure using expert opinions.

We have constructed the network by using an already existing DB where the available information are the risks with their frequencies, impacts and correlation among them. In total there are about 300 risks.

In our work we have considered only the frequencies of risks and no impacts. With our BN we construct the risks' joint probability and the impacts could be used in a later phase of scenario analysis to evaluate the loss distribution under the different scenarios [13].

In [Table 1](#) there is the DB structure used for network learning and in which each risk is considered as a binary variable (one if the risk exists (*yes*) and zero if the risk does not exist (*not*)). Therefore, for each considered risk in the network there will be one node with two states (*one* $\equiv Y$ and *zero* $\equiv N$).

So the task is to find the conditional probabilities tables by using only the correlations and the marginal frequencies. Instead, the net structure is obtained from [Table 1](#) by following the node relationships given by correlations.

The main ideas for finding a way to construct a BN have been: first to find the joint probabilities as functions of only the correlations and the marginal probabilities; second to understand how the correlations are linked with the incremental ratios or the derivatives of the child's probabilities as functions of the parent's probabilities. This choice is due to the fact that parent and child interact through the values of conditional probabilities; the derivatives are directly linked to such probabilities and, therefore, to the degree of interaction between the two nodes and hence with the correlation.

Table 1
Expert values database structure (learning table)

Parent	Child	Correlation	Parent freq.	Child freq.
Risk A	Risk B	$\rho_{AB} = 0.5$	$P(\text{risk A} = \text{Yes}) = 0.85$	$P(\text{risk B} = \text{Yes}) = 0.35$
Risk A	Risk C	$\rho_{AC} = 0.3$	$P(\text{risk A} = \text{Yes}) = 0.85$	$P(\text{risk C} = \text{Yes}) = 0.55$

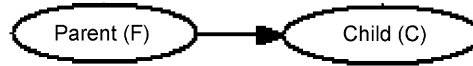


Fig. 1. One parent scheme.

Table 2
Equation systems for one parent scheme

CPT equation system		Joint equation system	
$\alpha_1 x + \alpha_2(1 - x) = y$	(1)	$c_1 = \rho_{FC}M + xy$	(1)
$\alpha_1 - \alpha_2 = k$	(2)	$c_2 = y - \rho_{FC}M - xy$	(2)
$\alpha_1 + \alpha_3 = 1$	(3)	$c_3 = x - \rho_{FC}M - xy$	(3)
$\alpha_2 + \alpha_4 = 1$	(4)	$c_4 = 1 - y - x + \rho_{FC}M + xy$	(4)

Once we have understood as to create equations, for the case with dependent parents we have used the local network topology to set the equations.

We have been able to calculate the Conditional Probability Table (CPT) up to three parents for each child. Although there is the possibility to generalize to more than three parents, it is necessary to have more data which are not available in our DB. So when four or more parents are present we have decided to divide and reduce to cases with no more than three parents. To approximate the network we have “separated” the nodes that give the same effects on the child (as for example the same correlations) by using auxiliary nodes [2]. When there was more than one possible scheme available we have used the mutual information (MI) criterion as a discriminating index by selecting the approximation with the highest total MI; this is the same to choose the structure with the minimum distance between the network and the target distribution [15,16].

We have analyzed first the case with only one parent to understand the framework, then it has been seen what happens with two independent parents and then dependent. Finally we have used the analogies between the cases with one and two parents for setting the equations for three parents.

2.1. One parent case solution

The case with one parent (Fig. 1) is the simplest. Let $P(F)$ and $P(C)$ be the marginal probability given from experts (as in Table 1):

- for the parent, F , we have: $P(F = Y) = x, P(F = N) = 1 - x$;
- for the child, C , we have: $P(C = Y) = y, P(C = N) = 1 - y$;

The equations to find either the conditional probabilities or the joint probabilities are in Table 2, where $k = \rho_{FC}\sqrt{\text{Var}[C]/\text{Var}[F]}$, ρ_{FC} is the correlation between nodes; $M = \sqrt{x(1-x)y(1-y)}$ and with α_i and c_i we indicate, respectively, the conditional and the joint probabilities.

Considering that probabilities c_i and α_i must be positive either the marginal probabilities or the correlation value should be constrained. If the marginal probabilities are fixed then the correlation values must be

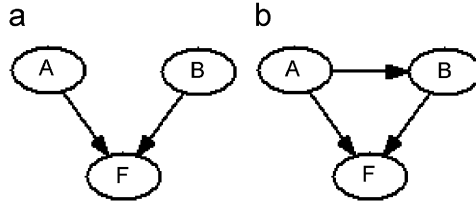


Fig. 2. Two independent parents (a) and dependent parents (b).

constrained, which will be normally the case, as estimates of probabilities are more reliable. It is not possible to have any value of correlation given the marginal probabilities. Indeed, as we want to maintain the marginal probabilities as fixed by the expert, correlation limits can be determined as follows:

$$\rho_{FC} > -\frac{xy}{M} = A, \quad \rho_{FC} > \frac{y + x(1-y) - 1}{M} = D, \quad \rho_{FC} < \frac{x(1-y)}{M} = B, \quad \rho_{FC} < \frac{y(1-x)}{M} = C;$$

and the correlation interval will be

$$\rho_{FC} \in [\max(A, D); \min(B, C)];$$

2.2. Two parents case solutions

This case (Fig. 2) is more complicated than the one before. In this situation a further difficulty is that the given expert correlations are only pairwise marginal and, therefore, we need more information to find the CPT.

For example the joint moment among the nodes which is not in the DB. Consequently, there can be more than one CPT, corresponding to different values of the joint moment, for the same marginal correlation and probability.

The joint moment becomes thus a project parameter to be set by using an appropriate criterion. We define the standardized joint moment among three variables to be

$$\rho_{N_i N_j N_k} = \frac{E[(N_i - E[N_i])(N_j - E[N_j])(N_k - E[N_k])]}{\sqrt[3]{\text{Var}[N_i]\text{Var}[N_j]\text{Var}[N_k]}}$$

where $E[\bullet]$ is the mean value and $\text{Var}[\bullet]$ is the variance ($E[(x - E[x])^2]$).

To choose among such CPTs we have used the total MI (I_{total}) by selecting the CPT with the $\rho_{N_i N_j N_k}$ that gives the minimum I_{total} .

In this case we have to distinguish between independent and dependent parents. The solutions are in Table 3, where $M_{BF} = \sqrt{z(1-z)y(1-y)}$, $M_{AF} = \sqrt{x(1-x)y(1-y)}$, $M_{AB} = \sqrt{x(1-x)z(1-z)}$ and $M_{ABF} = \sqrt[3]{x(1-x)z(1-z)y(1-y)}$. As before the α_i and c_i are, respectively, the conditional and the joint probabilities. The values x , y and z are the marginal probabilities defined in Table 1 by experts.

The problem is now setting the marginal correlations when those given from experts are not consistent with the marginal probabilities. Differently from the case with one parent where the correlation belongs to an interval, with two parents the admissible pairs (ρ_{AF}, ρ_{BF}) can be shown to belong to an area.

To approach this problem we have decided to decrease the values of the two correlations ρ_{BF} and ρ_{AF} with a fixed step by maintaining their relative difference. At each step we verified the existence of a value of ρ_{ABF} which supports the new pair (ρ_{AF}, ρ_{BF}) . If it exists the process is stopped, otherwise it goes to the next step; and so on.

If the correlation ρ_{AB} is different from zero (dependent parents), we can set it in advance using the interval obtained for the case of one parent; afterward the ρ_{AB} 's value is used into the joint equation. Then we can work again only on the pair (ρ_{AF}, ρ_{BF}) by considering the same procedure for independent parents and selecting $\rho_{N_i N_j N_k}$.

Table 3
Equation systems for two parents scheme

CPT equation system for independent parents		Joint equation system for dependent parents	
$f(x, z) = (\alpha_1 - \alpha_2 - \alpha_3 + \alpha_4)xz + (\alpha_2 - \alpha_4)x + (\alpha_3 - \alpha_4)z + \alpha_4 = y$	(1)	$c_1 = \rho_{ABF}M_{ABF} + xyz + (\rho_{AF}M_{AF})z + (\rho_{BF}M_{BF})x + (\rho_{AB}M_{AB})y$	(1)
$\frac{\partial f}{\partial x} = (\alpha_1 - \alpha_2 - \alpha_3 + \alpha_4)z + (\alpha_2 - \alpha_4) = \frac{(\rho_{AF})(M_{AF})}{x(1-x)}$	(2)	$c_1 + c_5 = \rho_{BF}M_{BF} + zy$	(2)
$\frac{\partial f}{\partial z} = (\alpha_1 - \alpha_2 - \alpha_3 + \alpha_4)x + (\alpha_3 - \alpha_4) = \frac{(\rho_{BF})(M_{BF})}{z(1-z)}$	(3)	$c_1 + c_3 = \rho_{AF}M_{AF} + xy$	(3)
$\frac{\partial^2 f}{\partial x \partial z} = \frac{\partial^2 f}{\partial z \partial x} = (\alpha_1 - \alpha_2 - \alpha_3 + \alpha_4) = \frac{(\rho_{ABF})(M_{ABF})}{x(1-x)z(1-z)}$	(4)	$c_1 + c_2 = \rho_{AB}M_{AB} + xz$	(4)
$\alpha_1 + \alpha_5 = 1$	(5)	$c_1 + c_2 + c_3 + c_4 = x$	(5)
$\alpha_2 + \alpha_6 = 1$	(6)	$c_1 + c_5 + c_6 + c_2 = z$	(6)
$\alpha_3 + \alpha_7 = 1$	(7)	$c_1 + c_3 + c_5 + c_7 = y$	(7)
$\alpha_4 + \alpha_8 = 1$	(8)	$c_8 = 1 - \sum_{i=1}^7 c_i$	(8)

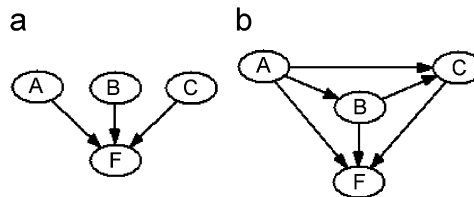


Fig. 3. Three independent parents (a) and dependent parents (b).

2.3. Three parents case solutions

As before, two equation systems are obtained. One system for the case with independent (see Fig. 3a) parents by which the CPT is directly calculated; another one when there are some correlations between parents and in this case the joint probabilities are calculated instead of the conditional ones (see Fig. 3b). To define the equation systems the analogies between cases with one and two parents have been exploited.

The solutions for independent and dependent parents are shown in Table 4. In such equations, obviously, there will be more missing data which are all the standardized joint moments among every two parents and the child and among all parents and the child. So what we do in such a situation is to use the procedure for the case of two parents for each pair of nodes and set the correlation values such that they will be feasible for all couples. Note that the correlation levels are now less than in previous cases. Moreover, in this case the standardized joint moment among all variables is set at zero to make the research less complex.

Furthermore, difficulties arise when there are large differences among the parents' marginal probabilities. Therefore, when there are more than three parents, we have decided to split them. Parents are split from the others by using the MI criterion [15,16].

As before, for the case of dependent parents to select the feasible marginal correlations and the standardized joint moments, we start to look for the admissible correlation between the nodes with one parent (A and B), then for the nodes with two parents (C has B and A as predecessor) and finally we set the joint moment and marginal correlations for the node with three parents (F). Obviously, now the procedure is more complex and it is more difficult to select the parameters.

3. Conclusion

So far we have seen that using the equation systems for conditional and joint probabilities the CPTs can be obtained. The method can be generalized to the case with more three parents, but there are problems in setting more parameters (standardized joint moment) and in looking for more complicated feasible marginal correlation areas.

Table 4
Equation systems for three parents scheme

CPT equation system for independent parents	Joint equation system for dependent parents
$f(x, z, w) = (\alpha_1 - \alpha_2 - \alpha_3 + \alpha_4 - \alpha_5 + \alpha_6 + \alpha_7 - \alpha_8)xzW$ $+ (\alpha_2 - \alpha_4 - \alpha_6 + \alpha_8)xz + (\alpha_3 - \alpha_4 - \alpha_7 + \alpha_8)wz$ $+ (\alpha_5 - \alpha_6 - \alpha_7 + \alpha_8)wz + (\alpha_6 - \alpha_8)z + (\alpha_4 - \alpha_8)x$ $+ (\alpha_7 - \alpha_8)w + \alpha_8 = y$	$c_1 = \rho_{ABCF}M_{ABCF} + xyzw + \rho_{ABC}M_{ABCY} + \rho_{ABF}M_{ABFW}$ $+ \rho_{ACF}M_{ACFZ} + \rho_{BCF}M_{BCFX} + \rho_{AB}M_{ABWY}$ $+ \rho_{AC}M_{ACZY} + \rho_{BC}M_{BCXY} + \rho_{AF}M_{AFXY}$ $+ \rho_{BF}M_{BFXW} + \rho_{CF}M_{CFXZ}$
$\frac{\partial f}{\partial x} = \frac{(\rho_{AF})\sqrt{x(1-x)y(1-y)}}{x(1-x)}$	$c_1 + c_2 + c_3 + c_4 + c_5 + c_6 + c_7 + c_8 = w$
$\frac{\partial f}{\partial z} = \frac{(\rho_{BF})\sqrt{z(1-z)y(1-y)}}{z(1-z)}$	$c_1 + c_2 + c_3 + c_4 + c_9 + c_{10} + c_{11} + c_{12} = x$
$\frac{\partial f}{\partial w} = \frac{(\rho_{CF})\sqrt{w(1-w)y(1-y)}}{w(1-w)}$	$c_1 + c_2 + c_5 + c_6 + c_9 + c_{10} + c_{13} + c_{14} = z$
$\frac{\partial^2 f}{\partial x \partial z} = \frac{(\rho_{ABF})\sqrt[3]{x(1-x)z(1-z)y(1-y)}}{x(1-x)z(1-z)}$	$c_1 + c_3 + c_5 + c_7 + c_9 + c_{11} + c_{13} + c_{15} = y$
$\frac{\partial^2 f}{\partial w \partial z} = \frac{(\rho_{ABF})\sqrt[3]{w(1-w)z(1-z)y(1-y)}}{w(1-w)z(1-z)}$	$c_1 + c_2 = \rho_{ABC}M_{ABC} + xyz + \rho_{AB}M_{ABW} + \rho_{AC}M_{ACZ} + \rho_{BC}M_{BCX}$
$\frac{\partial^2 f}{\partial x \partial w} = \frac{(\rho_{ABF})\sqrt[3]{w(1-w)x(1-x)y(1-y)}}{x(1-x)w(1-w)}$	$c_1 + c_9 = \rho_{ABF}M_{ABF} + xyz + \rho_{AF}M_{AFZ} + \rho_{BF}M_{BFX} + \rho_{AB}M_{ABY}$
$\frac{\partial^3 f}{\partial x \partial w \partial z} = \frac{(\rho_{ABCF})\sqrt[4]{w(1-w)z(1-z)y(1-y)x(1-x)}}{x(1-x)z(1-z)w(1-w)}$	$c_1 + c_3 = \rho_{ACF}M_{ACF} + xyz + \rho_{AF}M_{AFW} + \rho_{CF}M_{CFX} + \rho_{AC}M_{ACY}$
$\alpha_1 + \alpha_9 = 1$	$c_1 + c_5 = \rho_{BCF}M_{BCF} + xyz + \rho_{BF}M_{BFW} + \rho_{CF}M_{CFZ} + \rho_{BC}M_{BCY}$
$\alpha_2 + \alpha_{10} = 1$	$c_1 + c_2 + c_9 + c_{10} = \rho_{AB}M_{AB} + xz$
$\alpha_3 + \alpha_{11} = 1$	$c_1 + c_2 + c_3 + c_4 = \rho_{AC}M_{AC} + xw$
$\alpha_4 + \alpha_{12} = 1$	$c_1 + c_3 + c_9 + c_{11} = \rho_{AF}M_{AF} + xy$
$\alpha_5 + \alpha_{13} = 1$	$c_1 + c_2 + c_5 + c_6 = \rho_{BC}M_{BC} + zw$
$\alpha_6 + \alpha_{14} = 1$	$c_1 + c_3 + c_5 + c_7 = \rho_{CF}M_{CF} + wy$
$\alpha_7 + \alpha_{15} = 1$	$c_1 + c_5 + c_9 + c_{13} = \rho_{BF}M_{BF} + zy$
$\alpha_8 + \alpha_{16} = 1$	$c_{16} = 1 - \sum_i^{15} c_i$

So to develop a network we propose to use, separately, firstly the equations and procedure for the one parent; secondly those for two parents distinguishing when they are dependent and not. Finally we use the equations and the procedures, when possible, for the three parents case by distinguishing also in this situation between dependent and independent parents; otherwise we split one parent from the others by using the MI as splitting index [15,16].

We remark that we need to reduce to a more simple case those configurations with more than three parents. We can achieve this trying to estimate a local approximate structure, with only one, two and three parents, by “separating” those that give different effects on the child (as for instance different incremental ratios). If there are more schemes available for the substitution we select that with the highest MI (I_{total}) [15,16].

It is important to be noted that such method is modular, this is if we add or delete a node we can use the appropriate system (one, two or three nodes system equations) to according to we add or delete a parent or a child.

Acknowledgments

The authors acknowledge financial support from national Italian grant MIUR-FIRB 2006–2009 and European grant MUSING (contract number 027097).

References

- [1] R.G. Cowell, A.P. Dawid, S.L. Lauritzen, D.J. Spiegelhalter, *Probabilistic Networks and Expert Systems*, Springer, New York, USA, 1999.

- [2] F.V. Jensen, *Bayesian Networks and Decision Graphs*, Springer, New York, USA, 2001.
- [3] P. Giudici, *Applied Data Mining*, Wiley, West Sussex, England, 2003.
- [4] K.P. Murphy, An introduction to graphical models, May 10, from (<http://www.cs.ubc.ca>), 2001.
- [5] M.J. Druzdzel, L.C. van der Gaag, Building probabilistic networks: where do the numbers come from?, *IEEE Trans. Knowledge Data Eng.* 12 (4) (2000) 481–486.
- [6] T.S. Jaakkola, M.I. Jordan, Variational probabilistic inference and the QMR-DT database, *J. Artif. Intell. Res.* 10 (1999) 291–322.
- [7] D.A. Wiegmann, Developing a methodology for eliciting subjective probability estimates during expert evaluations of safety interventions: application for bayesian belief networks, Aviation Human Factors Division, October, from (<http://www.humanfactors.uiuc.edu>), 2005.
- [8] A.R. Daneshkhan, Uncertainty in probabilistic risk assessment: a review, The University of Sheffield, August 9, from (<http://www.shef.ac.uk/beep/publications.html>), 2004.
- [9] M.J. Druzdzel, L.C. van der Gaag, Elicitation of probabilities for belief networks: combining qualitative and quantitative information, in: *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, 1995, pp. 141–148.
- [10] A. Zagorecki, M. Druzdzel, An empirical study of probability elicitation under noisy-OR assumption, in: *Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference (FLAIRS 2004)*, 2004, pp. 800–885.
- [11] J.D. Francisco, F.G. Severino, Efficient Computation for the Noisy MAX, *Int. J. Intell. Systems* 18 (2003) 165–177.
- [12] O. Agnieszka, M. Druzdzel, H. Wasyluk, Learning Bayesian network parameters from small data sets: application of Noisy-OR gates, *Int. J. Approximate Reasoning* 27 (2001) 165–182.
- [13] C. Alexander (Ed.), *Operational Risk, Regulation Analysis and Management*, Prentice-Hall, London, England, 2003.
- [14] M.G. Cruz, *Modeling, Measuring and Hedging Operational Risk*, Wiley, West Sussex, England, 2002.
- [15] J. Williamson, Approximating discrete probability distributions with Bayesian networks, *IEEE Trans. Inform. Theory* 14 (3) (2000) 462–467.
- [16] G.D. Kleiter, R. Jirousek, Learning Bayesian networks under the control of mutual information, in: *Proceedings in Information Processing and Management of Uncertainty in Knowledge-Based Systems*, 1996, pp. 985–990.