Bayesian networks for transport decision scenarios

Alexander Holland
Department of Computer Science
University of Dortmund
44221 Dortmund (Germany)
Alexander.Holland@udo.edu

Abstract

Bayesian networks are formal graphical languages for representation and communication of decision scenarios requiring reasoning under uncertainty. We will analyze Bayesian networks and outline their advantages and disadvantages. Based on these assumptions we discuss transport decision scenarios under uncertainty. A transport planning approach like the postal delivery demonstrates a good framework for building and handling with normative systems like Bayesian networks.

Keywords: Bayesian networks, probabilities, uncertainties, transport planning.

1 Introduction

In this chapter we will give an introduction to the basic probability calculus, which is needed for understanding the technology of Bayesian networks, which will be discussed more precisely in the next chapter. First we will outline the basic axioms of probability [1]. Let Ω denote a finite collection of mutually exclusive statements about the world. Then by ε we denote the set of all events. The empty set Ø is the impossible event, the set Ω is called the certain event. In sequel we assume ε to be the power set of Ω, where Ω is a finite set . The pair *(Ω,ε)* is called the sample space. We define a function $P:\varepsilon \rightarrow [0,1]$ to be a probability function if it satisfies the following three conditions which are known as the Kolmogorov axioms:

(i) $P(A) \geq 0$ *for all* $A \subseteq \Omega$,
(ii) $P(\Omega) = 1$,
(iii) For $A, B \subseteq \Omega$ *from* $A \cap B \neq \emptyset$ *follows*
$P(A \cup B) = \mathrm{P(A)} + \mathrm{P(B)}$.

ε is a σ-field, i.e. a non-empty class of subsets of Ω that is closed under the formation of countable unions, countable intersections, complements and contains the sets Ø and Ω. Each probability satisfying the conditions *(i)-(iii)* has various properties like monotonicity, subadditivity or subtractivity. The notation *P(A)* stands for the unconditional or prior probability of all states $a_1, \ldots, a_n$ of the variable *A*, which are mutually exclusive. The prior probability for *A* is then

$$P(A) = (x_1, \ldots, x_n);\ x \geq 0;\ \sum_{i=1}^{n} x_i = 1, \tag{1}$$

where x_i is the probability of *A* being in state a_i. After retrieving evidence the prior probabilities are no longer applicable. Now posterior or conditional probabilities are used. Let *A* and *B* be any two events such that $P(B) \neq 0$. The conditional probability of *A* given *B* is defined by $P(A|B) = P(A \cap B)/P(B)$. Given the conditional probability we can introduce the Bayes rule. The Bayes rule tells how the posterior replaces the prior probability after receiving evidence, i.e. the observed data. We can express the Bayes rule as

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}. \tag{2}$$

Let $A_1,\ldots,A_n$ be a partition of Ω and let B be an event, such that $P(B) > 0$ and $P(A_i) > 0$ for all i, $i=1,\ldots n$. Then we can express $P(A_i|B)$ as

$$P(A_i \mid B) = \frac{P(B \mid A_i)P(A_i)}{\sum_{i=1}^{n} P(B \mid A_i)P(A_i)}. \tag{3}$$

2 Bayesian networks

The use of probabilistic models based on directed acyclic graphs apply within the field of artificial intelligence. Such models are known as Bayesian networks [2]. Their development was motivated by the need to model the top-down semantic and bottom-up perceptual combination of evidence in reading. The capability for bi-directional inferences, combined with a rigorous probabilistic foundation, were the reason for the appearance of Bayesian networks as a method of choice for reasoning under uncertainty in artificial intelligence and expert systems. A Bayesian network can be described as a graphical model for probabilistic relationships among a set of variables. It is therefore a graph in which the following holds:

- The nodes of the network represent a set of variables. The variables can be described as propositional variables of interest. Each variable has a set of finite mutually exclusive states.
- Pairs of the nodes are connected with a set of directed links. A link represents informational or causal dependencies among the variables.
- Each node has a conditional probability table $P(A|B_1,\ldots,B_n)$ attached that quantifies the effects that the parents $B_1,\ldots,B_n$ have on the node. We could say that the conditional probabilities encode the strength of dependencies among the variables.

In fact graphical models describe the distribution of a large number of random variables simultaneously. We can now represent a Bayesian network as n-dimensional discrete random variable $X_1,\ldots,X_n$. Each random variable X_k has the range $x_{k1},\ldots,x_{km_k} \in \Re$. We can define the conditional probability of X_k given C in the following:

$$p(x_{1j_1},\ldots,x_{nj_n} \mid C) = P(X_1 = x_{1j_1},\ldots,X_n = x_{nj_n} \mid C)$$
$$\text{with } j_k \in \{1,\ldots,m_k\}, k \in \{1,..n\} \tag{4}$$

or in compact form as

$$p(x_1,\ldots,x_n \mid C) = P(X_1 = x_1,\ldots,X_n = x_n \mid C). \tag{5}$$

Applying the chain rule we can also express (5) as

$$p(x_1,x_2,\ldots,x_n \mid C) = p(x_n \mid x_1,x_2,\ldots,x_{n-1},C) \cdot$$
$$p(x_{n-1} \mid x_1,x_2,\ldots,x_{n-2},C) \cdots p(x_2 \mid x_1,C)p(x_1 \mid C). \tag{6}$$

Given the direct acyclic graph (Figure 1) regarding the discrete random variables $X_1,\ldots,X_6$ the conditional probability follows applying formula (6):

$$p(x_1,\ldots,x_6 \mid C) = p(x_6 \mid x_4,C)p(x_5 \mid x_1,x_2,x_3,x_4,C)$$
$$p(x_4 \mid x_2,C)p(x_3 \mid x_1,x_2,C)p(x_2 \mid C)p(x_1 \mid C).$$

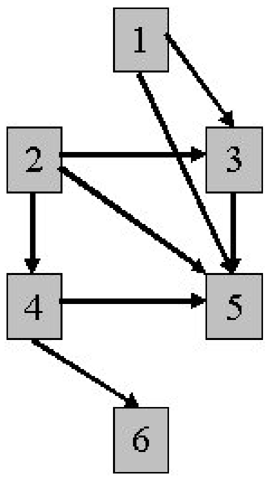

Figure 1: Bayesian network for postal delivery

3 Transport decision scenarios

Different logistic problems like planning problems deal with assignation from a number of sources to a number of destinations. Each source often offers amounts of goods, while the destination demands quantities of these goods. The planner wants to find the cheapest transporting schedule that satisfies the demand without violating the supply restraints. Many different transport problems can be formulated on networks [5]. Networks can be used to model many different applications such as vehicle routing problems or production planning. Consider for instance another application, a transport planning problem address the transport flow approach or transshipping scenario. In this paper we present a postal delivery approach embedded within the transport decision scenario. Regarding decision scenarios under uncertainty (see Figure 1) we can discuss the following task. A postal delivery organization supply packages. Several employees deliver theses packages depend on influencing factors like weather, package weights, traffic conditions or means of transport. Table 1 demonstrates the described scenario using six discrete random variables.

Table 1: transport random variables

variable
X_1 (package weight)
X_2 (weather)
X_3 (employee)
X_4 (traffic conditions)
X_5 (means of transport)
X_6 (information concern traffic conditions)

Traditionally, this problem has been analysed in two separate decisions. First, find out how many goods are supplied from each source to each destination and secondly what is the route that the distribution vehicle must follow in order to minimize the distribution costs. We discuss to determine the means of transport based on preliminary information. Figure 1 reflecting dependences and independences in the transport Bayesian network. The means of transport like heavy goods vehicles are dependent on X_1, X_2, X_3 and X_4. To determine a decision for instance X_i we calculate the probability of the values x_i of X_i. This means the summarization of the conditional probabilities in the following sense:

$$p(x_i \mid C) = \sum_{x_1} \cdots \sum_{x_{i-1}} \sum_{x_{i+1}} \cdots \sum_{x_n} p(x_1,\ldots,x_{i-1},x_i,x_{i+1},\ldots,x_n \mid C). \quad (7)$$

The information concerning the Bayesian network assessed by initializing random variables. We allocate the random variable X_k $(i < k < n)$ with the value $x_{ko} \in \{x_{k1},\ldots,x_{kmk}\}$. Applying the standardization constant α obtain

$$\begin{aligned} p(x_i \mid C) = \alpha \sum_{x_1} p(x_1 \mid C) \sum_{x2} p(x_2 \mid x_1, C) \cdots \\ \sum_{x_{i-1}} p(x_{i-1} \mid x_1,\ldots,x_{i-2},C) p(x_i \mid x_1,\ldots,x_{i-1},C) \\ \sum_{x_{i+1}} p(x_{i+1} \mid x_1,\ldots,x_i,C) \\ \cdots \sum_{x_{k-1}} p(x_{k-1} \mid x_1,\ldots,x_{k-2},C) p(x_{k0} \mid x_1,\ldots,x_{k-1},C). \end{aligned} \quad (8)$$

Using the likelihood function $\lambda(x_1,\ldots x_i)$ substitute

$$\begin{aligned} \lambda(x_1,\ldots,x_{k-1}) &= p(x_{k0} | x_1,\ldots,x_{k-1},C) \\ \lambda(x_1,\ldots,x_{k-2}) &= \\ \sum_{x_{k-1}} p(x_{k-1} | x_1,\ldots,x_{k-2},C) \lambda(x_1,\ldots,x_{k-1}) \\ \cdots \\ \lambda(x_1,\ldots,x_i) &= \\ \sum_{x_{i+1}} p(x_{i+1} | x_1,\ldots,x_i,C) \lambda(x_1,\ldots,x_{i+1}) \end{aligned} \quad (9)$$

we obtain from (8)

$$\begin{aligned} p(x_i \mid C) = \alpha \sum_{x_1} p(x_1 \mid C) \sum_{x2} p(x_2 \mid x_1, C) \cdots \\ \sum_{x_{i-1}} p(x_{i-1} \mid x_1,\ldots,x_{i-2},C) p(x_i \mid x_1,\ldots,x_{i-1},C) \cdot \\ \lambda(x_1,\ldots,x_i). \end{aligned} \quad (10)$$

In the following illustrated Figure 2 we can see the dependences regarding the four random variables weather, feeling of the employee, means of transport and pre-information concerning the means of transport.

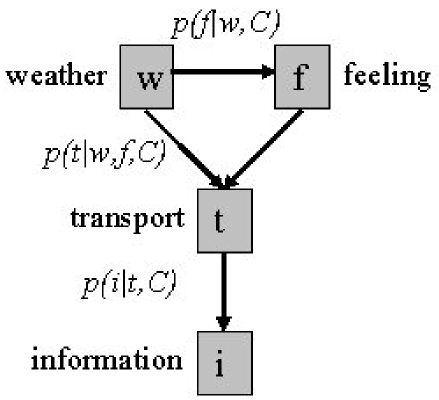

Figure 2: Bayesian network transport scenario

To determine the choice for the means of transport we calculate $p(t|C)$ as follows

$$p(t \mid C) = \alpha \sum_{w} p(w \mid C) \sum_{f} p(f \mid w, C) p(t \mid w, f, C) p(i \mid t, C). \quad (11)$$

Algorithms for the computation of the posterior probabilities with discrete variables using conditional independence properties was developed and for instance available at [3] or [4]. Choosing the means of transport based on the following preliminary data

$p(i \mid t, C) = 0{,}6$ *(hgv 1), 0,4 (hgv 2),*

$$p(w \mid C) = \begin{bmatrix} 0{,}3 & rainy \\ 0{,}5 & cloudy \\ 0{,}2 & sunny \end{bmatrix},$$

$$p(t \mid w, f_{rested}, C) = \begin{bmatrix} hgv1 & hgv2 & \\ 0{,}2 & 0{,}8 & rainy \\ 0{,}7 & 0{,}3 & cloudy \\ 0{,}9 & 0{,}1 & sunny \end{bmatrix},$$

$$p(t \mid w, f_{tired}, C) = \begin{bmatrix} hgv1 & hgv2 & \\ 0{,}1 & 0{,}9 & rainy \\ 0{,}6 & 0{,}4 & cloudy \\ 0{,}8 & 0{,}2 & sunny \end{bmatrix},$$

$$p(f \mid w, C) = \begin{bmatrix} rested & tired & \\ 0{,}2 & 0{,}8 & rainy \\ 0{,}4 & 0{,}6 & cloudy \\ 0{,}7 & 0{,}3 & sunny \end{bmatrix},$$

we obtain

p(t|C) = 0,63 (hgv 1) and *p(t|C) = 0,37 (hgv 2)*

preferring the first heavy good vehicle. Choose another assumption instantiate by *w=cloudy* derive the probability *p(t|w=cloudy,C) = 0,73 (hgv 1)* and *p(t|w=cloudy,C)= 0,27 (hgv 2).* Assume also *f=tired* for the driver we obtain *p(t|f=tired,w=cloudy,C) = 0,69 (hgv 1)* and *p(t|f=tired,w=cloudy,C)= 0,31 (hgv 2).* The modified probabilities differ only marginal.

References

[1] S. Russel; P. Norvig (1995). *Artificial Intelligence – A modern approach*. Prentice Hall Inc., 1995.

[2] F.V. Jensen (2001). *Bayesian networks and decision graphs.* Statistics for engineering and information science. Springer, 2001.

[3] J. Pearl (1988). *Probabilistic reasoning in intelligent systems.* CA: Morgan Kauffman, 1988.

[4] B. Das (1999). *Representing uncertainties using Bayesian networks.* Electronics and Surveillance Research Laboratory, Australia, 1999.

[5] J. Heizer; B. Render (1996). *Production and operations management. Strategic and Tactical Decisions.* Prentice Hall, Fourth Edition, New Jersey, 1996.