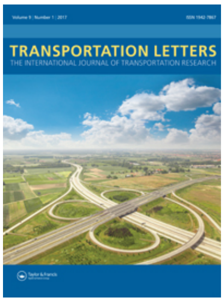


Transportation Letters

The International Journal of Transportation Research

ISSN: 1942-7867 (Print) 1942-7875 (Online) Journal homepage: www.tandfonline.com/journals/ytrl20

Road accident data analysis using Bayesian networks

Ali Karimnezhad & Fahimeh Moradi

Road accident data analysis using Bayesian networks

Ali Karimnezhad* and Fahimeh Moradi

Bayesian Networks (BNs) are graphical probabilistic models representing the joint probability function over a set of random variables using a directed acyclic graphical structure. In this paper, we consider a road accident data set collected at one of the popular highways in Iran. Implementing the well-known parents and children algorithm, as a constraint-based approach, we construct a BN model for the available accident data. Once the structure of the BN is learnt, we concentrate on the parameter-learning task. We compute the maximum-likelihood estimates of some parameters of interest, specifically, conditional probability of fatalities in the network.

Keywords: Bayesian networks, Directed acyclic graph, Maximum-likelihood estimation, PC algorithm, Road accident data, Structure learning

Introduction

Bayesian Networks (BNs) are one of the most popular models (Pearl, 1988; Jensen and Nielsen, 2007) presenting joint probability distributions of a given set of random variables. Learning BNs from data is normally split into two related steps: (1) learning the structure of the network and (2) learning the parameters (Heckerman *et al.*, 1995, 1999, Cheng *et al.*, 2002).

There are two main approaches to the structure learning in BNs. One approach, namely constraint-based approach, is based on carrying out several independence tests on the database and building a BN in agreement with tests results. The main example of this approach is the parents and children (PC) algorithm (Spirtes et al., 2000). The other approach, which is usually referred to as score-based approach, is based on some edge-based scoring and searching methods to evaluate a given BN model (Heckerman *et al.*, 1995). The desired BN with an optimal score is determined by searching over the space of all possible BN models.

Once the structure of a network is determined, parameter learning becomes possible. The literature abounds with different parameter learning methods, see Buntine (1991), Cooper and Herskovits (1991), Ramoni and Sebastiani (2003), de Campos and Qiang (2008) and Karimnezhad and Moradi (2015) among others. The parameter-learning task is usually done by maximizing the joint probability density functions of a set of given variables. The resulting estimates are referred to as maximum-likelihood (ML) estimates. The main reason for popularity of ML criterion is that it is based on a well-founded probabilistic principle which implies that the ML estimates often are optimal in a statistically meaningful sense. This optimality may be for finite sample sizes or asymptotically as the sample size tends to infinity, see Eggermont and LaRiccia (2001) for more information.

BNs have recently attracted attention in transportation engineering. The number of traffic accidents and their effects, mainly human fatalities and injuries, highlights the importance of analyzing the factors that contribute to their occurrence (de Oña *et al.*, 2011). To the best of our knowledge, the first application of BNs found its way to transportation engineering in 2003 in which Davis and Pei (2003) illustrated how BNs can be used to support inductive reasoning about road accident. Later, Simoncic (2004) used BNs to construct two car accident injury severity model. A BN was built using several variables, and the most probable explanation was calculated for the most probable configuration of values for all the variables in the BN, in order to serve as an indication of the quality of the estimated BN. Marsh and Bearfield (2004) used them for accident modeling on the UK railway network. Ozbay and Noyan (2006) applied BNs to investigate incident clearance duration time on road links. Janssens *et al.* (2006) criticized the use of decision trees and encountering advantages of BNs, they integrated BNs and decision trees to achieve a better predictive model. Gregoriades (2007) highlighted the interest of using BNs to model traffic accidents and emphasized on the need of modeling the uncertainties involved in the factors that can lead to road accidents. He listed a number of candidate approaches for modeling uncertainty of which Bayesian probability and BNs are prime examples. Based on data survey and statistical analysis, a BN for traffic accident causality analysis was developed by Xu *et al.* (2011). The structure and parameter of the BN were learnt using the K2 algorithm and Bayesian parameter estimation, respectively. Readers may refer to Hongguo *et al.* (2010), Wach (2013) and Deublein *et al.* (2014) for some related works and brief descriptions of advantages and disadvantages of BNs.

This paper has been organized as follows: We first deal with structure learning and introduce the well-known PC algorithm as a useful constraint-based approach. Afterward, we provide

School of Mathematics, Statistics and Computer Science, University of Tehran, Tehran, Iran

*Corresponding author, email a.karimnezhad@khayam.ut.ac.ir

Received 21 December 2014; accepted 02 December 2015
DOI 10.1080/19427867.2015.1131960

some prevalent techniques that are usually used for learning parameters of BNs. Implementing the PC algorithm, we construct a BN model for a set of real road accident data collected at Hamedan-Qazvin highway, one of the popular highways in Iran (March 2009–December 2013). After determining the appropriate BN structure, we compute the maximum likelihood estimates of some parameters of interest. We end this paper by providing some conclusions.

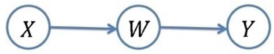
Learning BNs

When dealing with the BNs, presumably the most challenging task is learning their structure, and parameter learning stands in the second order. Because of enormous usefulness of learning the structure of BNs, it has been considered in many different subjects, such as biology, medicine, chemistry, physics, transportation engineering, etc.

In this section, we provide some preliminaries required for pursuing the sequel. In this regard, we first discuss on structure learning and introduce the PC algorithm. Then, we provide a brief review on parameter learning based on the ML approach.

Learning the structure

Perhaps the problem of learning the most probable BN structure from data is a challenging problem. To make the structure-learning task clear, we provide a brief explanation of the structure-learning procedure. Readers may refer to Jensen and Nielsen (2007) for a comprehensive explanation. Consider a situation in which some agent produces samples of cases from a BN $\mathcal{G}$ over the universe $\mathcal{U}$. The cases are handed over to an analyst, and she/he is asked to reconstruct the BN from the cases. This is the general setting for structural learning of BNs. Obviously, in the real world, the analyst may not be sure that the cases are actually sampled from a "true" network, but here, we assume this can be possible and the sample is fair. That is, the set $\mathcal{D}$ of cases reflects the distribution $P_{\mathcal{G}}(\mathcal{U})$ determined by $\mathcal{G}$. In other words, the distribution $P^*_{\mathcal{G}}(\mathcal{U})$ of the cases is very close to $P_{\mathcal{G}}(\mathcal{U})$. Furthermore, we assume that all links in $\mathcal{G}$ are essential, i.e., if one removes a link, then the resulting network cannot represent $P(\mathcal{U})$. The task is now to find a BN, $\mathcal{G}$, close to $\mathcal{G}$. In principle, this can be done by performing parameter learning for all possible structures and then selecting as candidates those models for which $P_{\hat{\mathcal{G}}}(\mathcal{U})$ is close to $P^*_{\mathcal{D}}(\mathcal{U})$. This very simple approach leads us to the most probable BN structure but we are faced with three problems, which are fundamental for learning BNs. First of all, the space of all BN structures is extremely large. It has been shown that the number of different structures grows more than exponentially in the number of nodes. Secondly, when searching through the network structures, we may end up with several equally good candidate structures. Since a BN over a complete graph can represent any distribution over its universe, we know that we will always have several candidates, but a BN over a complete graph will hardly be the correct answer. If so, it is a very disappointing answer. Thirdly, we have the problem of overfitting, i.e., the selected model is so close to $P^*_{\mathcal{D}}(\mathcal{U})$ that it also covers the smallest deviances from $P_{\mathcal{G}}(\mathcal{U})$. Again, a complete graph can represent $P^*_{\mathcal{D}}(\mathcal{U})$ exactly, but $\mathcal{D}$ may have been sampled from a sparse network. For some detailed discussion readers may refer to Jensen and Nielsen (2007). Dealing with learning the structure of BNs, two well-known approaches are followed basically; constraint-based and score-based approaches. The constraint-based methods establish a set of conditional independence statements holding for the data and use this set to build a network with (so-called) *d*-separation properties corresponding to the conditional independence properties determined. The score-based methods produce a series of candidate BNs, calculate a score for each candidate, and return a candidate of highest score. For more information, see Jensen and Nielsen (2007).



1 A typical three-node causal model

Constraint-based learning methods systematically check the data for conditional independence relationships. Typically, the algorithms run a χ^2-independence test when the data set is discrete and a Fisher's exact $\mathcal{Z}$ test when it is continuous in order to decide on dependence or independence, that is, upon the rejection or acceptance of the null hypothesis of conditional independence.

It is interesting to mention that although, the score-based procedure has some advantages in finding a BN structure (see Heckerman *et al.*, 1999), but some advantages of the constraint-based learning make us still willing to use them. While the score-based approach is efficient for learning the BN structure, the ability to scale up to hundreds of thousands of variables is a key advantage of the constraint-based approach against the alternative score-based one. Specially, the PC algorithm is appreciated in many contexts, see Jensen and Nielsen (2007) and de Morais (2009) among others. The constraint-based approach has an intuitive basis and under some ideal conditions, it has guarantee of recovering a graph equivalent to the one being a true model for the data. It can be considered as a smart selection and ordering of the questions that have to be done in order to recover a causal structure (Cano *et al.*, 2008). Keeping the advantages of using PC algorithm in mind, we are interested in learning the structure of the road accident network using the PC algorithm.

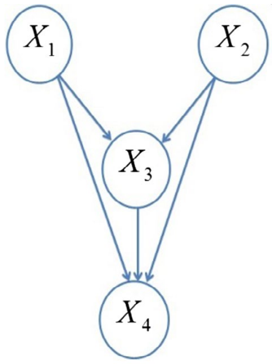
The PC algorithm

In this section, we briefly provide details of structure learning using the PC algorithm to help readers better understand the model construction procedure.

Consider the situation in Figure 1 in which the variable X has an influence on W, which in turn has an influence on Y. Obviously, evidence about X will influence the certainty of W, which then influences the certainty of Y. Similarly, evidence about Y will influence the certainty of X through W. On the other hand, if the state of W is known, then the channel is blocked, and X and Y become independent. In this case, we say that X and Y are *d*-separated given W (*d* stems from directed graph) and denote it by $I(X, Y, W)$. Similarly, $I(X, Y, \mathcal{W})$ means that X and Y are *d*-separated given any variable W in $\mathcal{W}$.

As an illustration of *d*-separation concept, Figure 2 shows a causal model for the relations between *Rainfall* (no, light, medium, heavy), *Water Level* (low, medium, high), and

2 A causal model for rainfall, water level, and flooding



3 A typical four-node BN

Flooding (yes, no). If we have not observed the water level, then knowing that there has been a flooding, will increase our belief that the water level is high, which in turn will tell us something about the rainfall. The same line of reasoning holds in the other direction. On the other hand, if we already know the water level, then knowing that there has been flooding will not tell us anything new about rainfall. Thus, we have $I(Rainfall, Flooding, Water\ Level)$, which means that *Rainfall* and *Flooding* are *d*-separated given *Water Level*, see Jensen and Nielsen (2007) for more details.

The PC algorithm is based on a two-step process. First, using the conditional independence tests, it determines the skeleton of a BN (it specifies some undirected edges between nodes). Then, at the second step, on the basis of the *d*-separation concept, it determines the direction between two nodes (for the two nodes X and Y, it determines whether $X \longrightarrow Y$ or $X \longleftarrow Y$ can be true).

The first step can be restated in terms of a repeated algorithm. To do so, let $|\mathcal{W}|$ and $\mathrm{nb}(X)$ denote number of elements in the set $\mathcal{W}$ and the set of neighbors of X, respectively. Then, the test sequence can be stated as below:

1. Start with the complete graph;
2. Take $i = 0$;
3. While a node has at least $i + 1$ neighbors;
   - for all nodes X with at least $i + 1$ neighbors;
     - for all neighbors Y of X;
       - for all neighbor sets $\mathcal{W}$ such that $|\mathcal{W}| = i$ and $\mathcal{W} \subseteq \mathrm{nb}(X)\backslash\{Y\}$;
         - if $I(X, Y, \mathcal{W})$, then remove the link $X - Y$ and store $I(X, Y, \mathcal{W})$;
   - Step up by 1, i.e. Take $i = i + 1$.

For a full discussion on performance of the PC algorithm and its usefulness, see Spirtes *et al.* (2000) and Jensen and Nielsen (2007). Application of the PC algorithm for determining a BN structure is considered in the next section.

Learning the parameters

In this subsection, we assume readers are familiar with BNs and related topics. However, for an easy pursuing matter, we briefly provide some fundamentals. For more information, see Jensen and Nielsen (2007) and Koski and Noble (2009).

A causal network consists of a set of variables and a set of directed links (also called arcs) between variables. Causal networks can be used to follow how a change of certainty in one variable may change the certainty for other variables. Mathematically, the structure is called a directed graph. The directed graph is called acyclic, if does not contain any directed cycle.

Now, consider a directed acyclic graph, DAG for short, $\mathcal{G} = (V, E)$ with $V = \{X_1, \ldots X_d\}$ denoting a set of variables and E standing for the set of edges contained in the Cartesian product $V \times V$. Suppose that for each $j = 1, \ldots, d$, the variable X_j takes values in the set $\mathcal{X}_j = \{x_j^1, \ldots, x_j^{k_j}\}$. Thus, the set of all possible outcomes for the experiment may be denoted by the following:

$$\mathcal{X} = \{\mathcal{X}_1 \times \ldots \times \mathcal{X}_d\} = \{x_1^{j_1}, \ldots, x_d^{j_d} | j_i = 1, \ldots, k_i, i = 1, \ldots, d\}.$$

Hence, a sample of cases is given by the following:

$$\boldsymbol{x} = \begin{pmatrix} \boldsymbol{x}_{(1)} \\ \vdots \\ \boldsymbol{x}_{(n)} \end{pmatrix}$$

where $\boldsymbol{x}_{(i)} = (x_{i,1}^{(j_1)}, \ldots, x_{i,d}^{(j_d)})$ denotes the i-th complete instantiation. For each variable X_j, consider all possible instantiations of the parent set Π_j and denote them by the set $\left\{\pi_j^{(1)}, \ldots, \pi_j^{(q_j)}\right\}$. Hence, $\pi_j^{(l)}$ implies that the parent configuration of variable X_j is in state $\pi_j^{(l)}$ and there are q_i possible configurations of Π_j.

For a given graph structure $\mathcal{G} = (V, E)$, let:

$$n_{jilk} = \begin{cases} 1, if\, (x_j^{(i)}, \pi_j^{(l)}) \text{ is found in } \boldsymbol{x}_{(k)} \\ 0, \text{otherwise,} \end{cases}$$

where $(x_j^{(i)}, \pi_j^{(l)})$ is a configuration of the family (X_j, Π_j). Let $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ denote the set of parameters defined by:

$$\theta_{ji,l} = P(X_j = x_j^{(i)} | \pi_j = \pi_j^{(l)}),$$

for $l = 1, \ldots, q_j, i = 1, \ldots, k_j, j = 1, \ldots, d$, with $\sum_{i=1}^{k_j} \theta_{ji,l} = 1$.

Using the decomposition of the probability distribution defined by the BN, the joint probability of a case $\boldsymbol{x}_{(k)}$ occurring may be written as:

$$p_{X_{(k)}|\boldsymbol{\Theta}}\left(\boldsymbol{x}_{(k)} | \theta, E\right) = \Pi_{j=1}^{d} \Pi_{l=1}^{q_j} \Pi_{i=1}^{k_j} \theta_{ji,l}^{n_{jilk}}.$$

For independent observations $(\boldsymbol{x}_{(1)}, \ldots, \boldsymbol{x}_{(n)})$, the joint probability of the cases is given by the following:

$$p_{X|\boldsymbol{\Theta}}(\boldsymbol{x}|\theta, \mathcal{G}) = \Pi_{k=1}^{n} \Pi_{j=1}^{d} \Pi_{l=1}^{q_j} \Pi_{i=1}^{k_j} \theta_{ji,l}^{n_{jilk}} = \Pi_{j=1}^{d} \Pi_{l=1}^{q_j} \Pi_{i=1}^{k_j} \theta_{ji,l}^{n_{jil.}}.$$

where $n_{jil.} = \sum_{k=1}^{n} n_{jilk} = 1$.

The likelihood function is then given by the following:

$$L(\boldsymbol{\theta}) = \Pi_{j=1}^{d} \Pi_{l=1}^{q_j} \Pi_{i=1}^{k_j} \theta_{ji,l}^{n_{jil.}}.$$

One can observe that the ML estimate of $\theta_{ji,l}$ is given by the following:

$$\delta_{ji,l}^{\mathrm{ML}} = \frac{n_{jil.}}{n_{j.l.}},$$

where $n_{j.l.} = \sum_{i=1}^{k_j} n_{jil.} = 1$.

Example 1. Consider the DAG depicted in Figure 3 with four binary nodes $X_1, \ldots, X_4$. Hence, $d = 4$, $k_i = 2$ for $i = 1, \ldots, 4$, and the parent set of X_3 has four possible instantiations:

$$\pi_3^{(1)} = (0,0), \pi_3^{(2)} = (0,1), \pi_3^{(3)} = (1,0), \pi_3^{(4)} = (1,1).$$

Suppose complete instantiations of five cases are available as below:

$$\boldsymbol{x} = \begin{pmatrix} (0,0,1,0) \\ (0,1,0,1) \\ (1,1,1,1) \\ (0,0,0,0) \\ (1,0,1,1) \end{pmatrix},$$

and assume we are interested in learning the parameter $\theta_{31,1} = P(X_3 = 0 | X_1 = 0, X_2 = 0)$. Then, the ML estimate of $\theta_{31,1}$ is given by $\delta_{31,1}^{\mathrm{ML}} = \frac{n_{311.}}{n_{3.1.}} = \frac{1}{2}$.

Implementing the PC algorithm in modeling the road accident data

In this section, we construct a BN model to analyze the available road accident data to illustrate practical utility of the PC algorithm. The available data have been collected at Hamedan-Qazvin highway, one of the popular highways in Iran (March 2009–December 2013).

Data description

The available road accident data show that 65,535 accidents have been recorded during a 56-month period. The data set consists of the following information of drivers in accident scenes:

(1) X_1; Sex (male, female),
(2) X_2; Education (unread, higher secondary, diploma, associate's degree, bachelor's degree, graduate degree, unknown),
(3) X_3; Vehicle type (pickup, motorcycle, bus, minibus, truck, trailer, car, other),
(4) X_4; Age (young, middle-aged, old, unknown),
(5) X_5; License type (type 1, type2, expired, conditional, special, not seen, without license),
(6) X_6; Seatbelt status (used, unused, unknown),
(7) X_7; Injury type (not injured, injured, death).

Except the *Age* variable, all the variables were discrete. The *Age* variable was discretized separately due to implementation of the PC algorithm.

Of the 65,535 total recorded accidents, the majority of drivers were male (about 90.4%). The recorded information of drivers' education shows that percentage of drivers with the education types unread, unknown, higher secondary, diploma, associate's degree, bachelor's degree and graduate degree was 1.10, 58.53, 1.29, 37.17, 0.61, 1.00 and 0.30, respectively. This reveals that drivers with highest educational degree, i.e., graduate degree, were involved in minority of the accidents. Vehicle type was reported in eight different groups. Percentage of accidents for the different eight types pickup, motor cycle, bus, minibus, truck, trailer and car was reported as 12.39, 10.1, 3.64, 2.10, 11.50, 1.86 and 57.80, respectively. The remaining 0.61% were categorized as others. 42.8% of drivers were middle-aged, while 6.20% and 35.3% were categorized as old and young drives. In the remaining 15.7% of the cases, driver age was not recorded. Driver license type was reported in seven different categories. Due to the rules in Iran, not every driving license holder can drive any vehicles. For example, type 2 license holders are not allowed to drive buses, trucks and trailers. Percentage of license holders with type 1, type 2, expired, conditional and special driving licenses was 15.21, 56.29, 0.012, 0.21, 0.048, respectively. In 2.73% of the cases, drivers had no driving license and in the remaining 25.5%, it was reported as unseen. 76.3% of drivers had worn their seat belt, while 8.5% of them had not used their seat belt. In the remaining 15.2% of the accidents, seat belt use was reported unknown. Of the 65,535 recorded accidents, 1.9% led to driver death and in 11.8% of the cases, injuries were reported. The remaining 86.3% of the accidents led to no significant injuries. All the percentages described here are summarized in Figure 5.

Structure learning for the road accident data

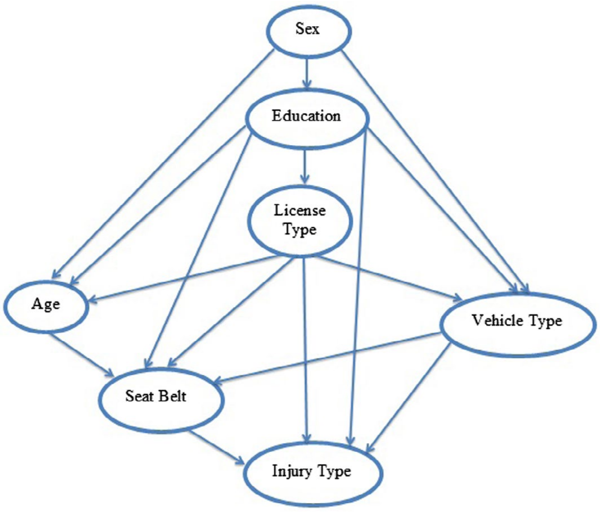
To determine the structure of the road accident network, we implemented the PC algorithm for the available dataset. The resulting structure is depicted in Figure 4.

It is interesting to realize from Figure 4 that, the variable *Injury Type* directly is influenced by the variables *Seat Belt*, *License Type*, *Vehicle Type* and *Education*. It is also indirectly affected by *Sex* and *Age*. Some other causal relations might be of interest. One might be interested in the chance of being injured while wearing seat belt when the *Vehicle Type* is bus. The question that *how much seat belt wearing can decrease the probability of being injured* (while driving a specific vehicle) might be of another interest. Finding such probabilities and quantities is the subject of the next subsection.

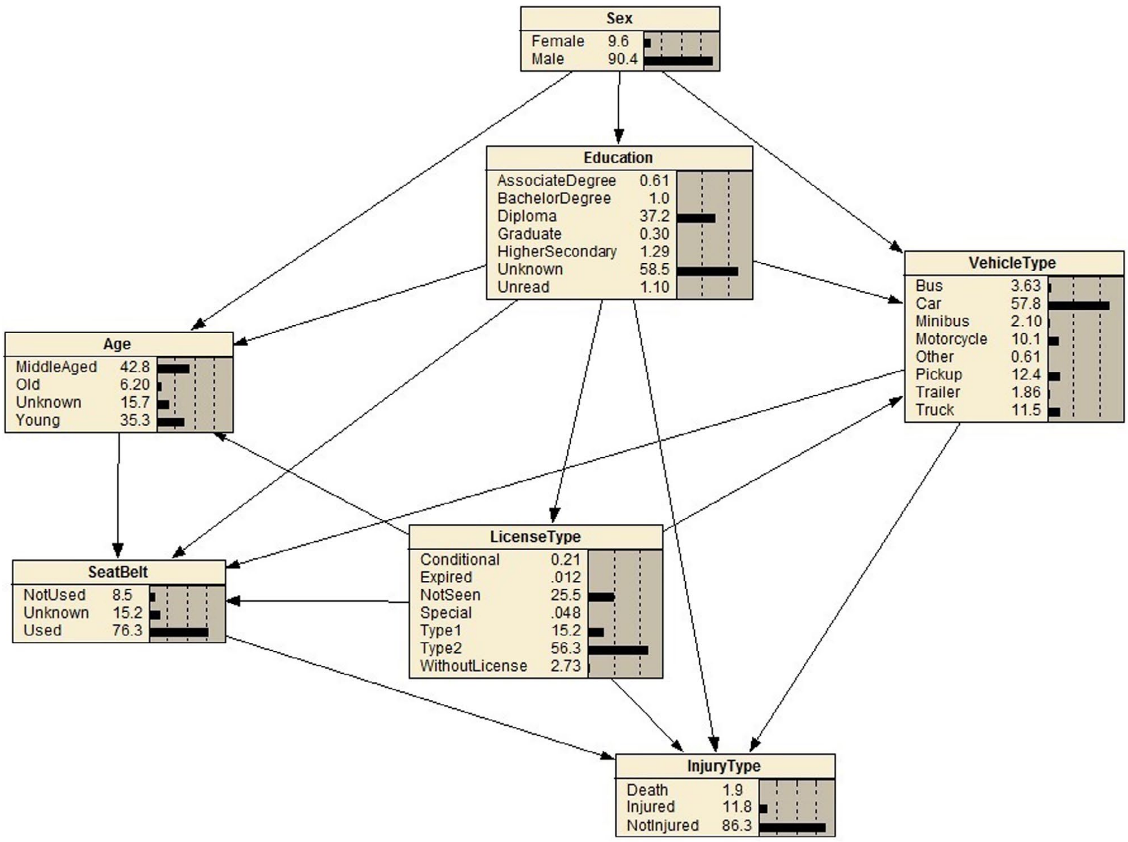
Parameter learning in the road accident network

In this section, we consider the road accident network shown in Figure 4 and our concentrate is on learning some interesting conditional probabilities (to decipher influence of the nodes on each other). Also, we provide some results of the ML parameter learning. It is worth noting the ML estimates usually have optimal properties for sufficiently large sample data. For this reason, we compute the ML estimates of some desired probabilities, as optimal estimates. For more information about optimality of the ML estimates, see Eggermont and LaRiccia (2001) and Millar (2011) among others.

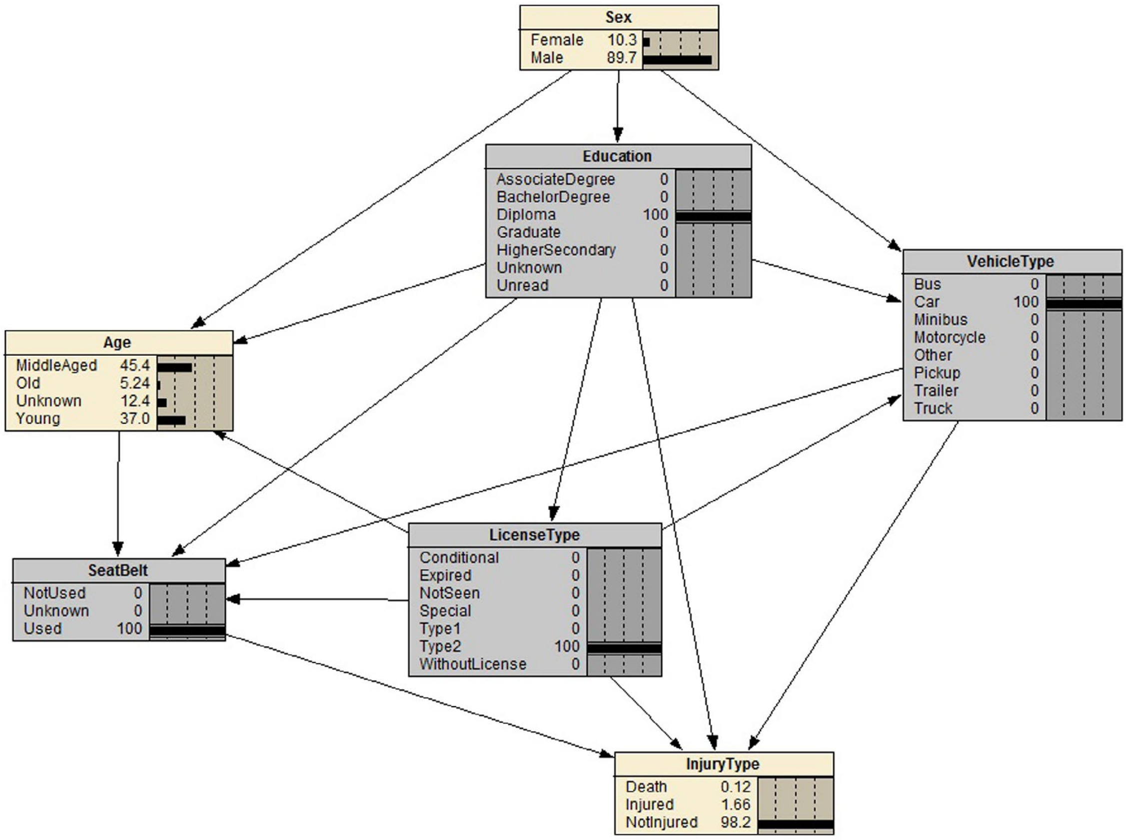
Before computing the conditional probabilities $\theta_{ji,l}$ for specific values of the j, i, and l, we would like to tabulate the unconditional probabilities in Figure 5, which has been provided using the *Netica* software. By the unconditional term, we mean that the reported probabilities are those which describe



4 The learnt road accident network



5 The road accident network



6 Conditional probabilities of fatalities while using seat belt in the road accident network

each of the nodes regardless the information of other nodes. As a special case, the probabilities in Figure 5 for the node *Sex*, illustrate that 9.6% of drivers were female and the remaining were male. As another case, the probabilities for the node *Seat Belt* reveal that 76.3% of drivers had worn their seat belts, and 8.50% of them had not worn their seat belts. For the remaining 15.2% of the accidents, seat belt status was recorded unknown.

To obtain ML estimates of the road accident network, let us first focus on computing a specific conditional probability: the chance of being injured while wearing seat belt and driving a car. Further, suppose the driver has a diploma degree and a type 2 driving license. We can mention this probability in terms of the notations introduced earlier. To do so, first notice that the set (X_2, X_3, X_5, X_6) makes up the parents of the node X_7 (*Injury Type*). Now, we need to assign some ordered numbers to each status of the variables. To make a systematic numbering, we assign 1 to the first status (row) in each box in Figure 5 and increase the numbers by one, when we jump one row down. We keep this type of numbering until all the statuses are numbered. Thus, as an example, for the variable X_1 (Sex in Figure 5), we assign 1 to female and 2 to male. So, the possible parent sets can be represented by $\pi_7^{(l)} = (i_1, i_2, i_3, i_4)$, where $i_1 = 1, \ldots, 7$, $i_2 = 1, \ldots, 8$, $i_3 = 1, \ldots, 7$ and $i_4 = 1, \ldots, 3$, $l = 1, \ldots, 7 \times 8 \times 7 \times 3$. Thus, for our problem, the special parent set would be $\pi_7^{(375)} = (3, 2, 6, 3)$ and the desired probability is $\theta_{72,375}$ which will be estimated by its ML estimate as below:

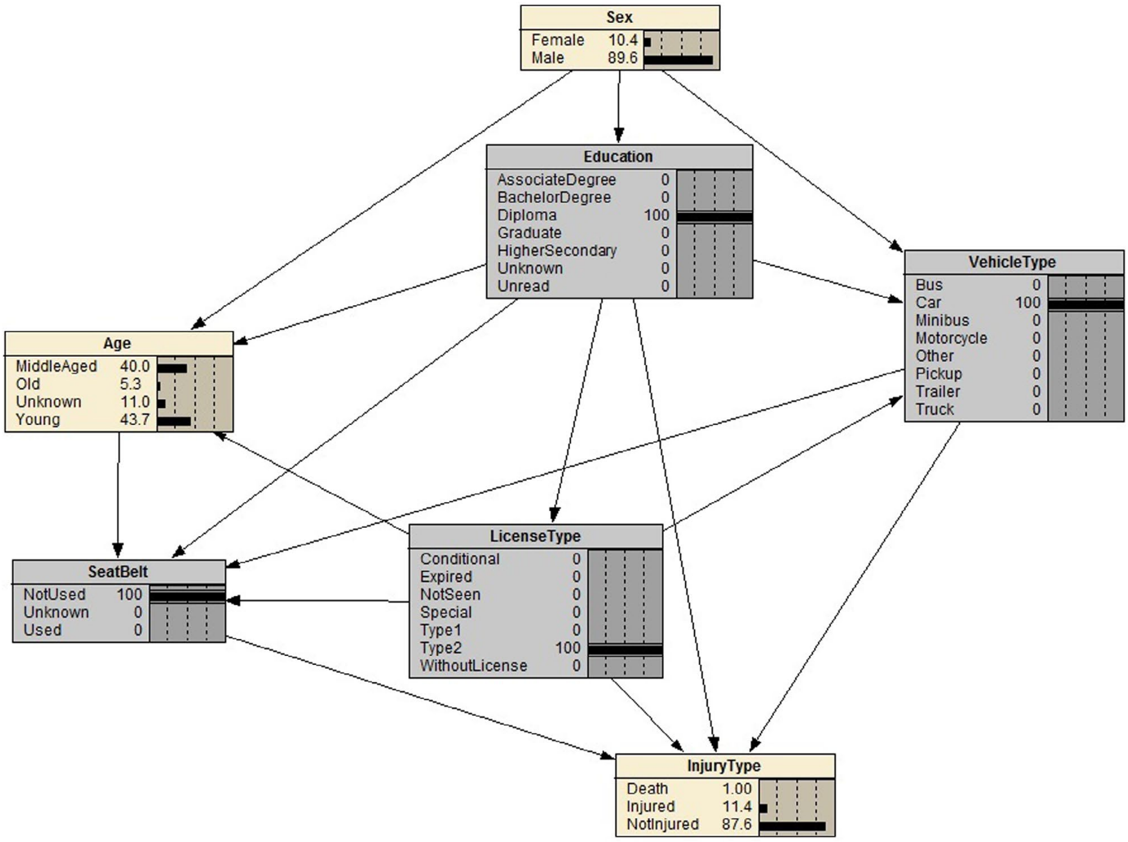
$$\delta^{\mathrm{ML}}_{72,375} = P\left(X_7 = x_7^{(2)} | \pi_7 = \pi_7^{(375)}\right) = P\left(X_7 = 2 | \pi_7 = (3, 2, 6, 3)\right) = 0.0166.$$

Thus, the chance of being injured while wearing seat belt and driving a car, knowing that the driver has a diploma degree and a type 2 driving licence is 1.66%.

One might also be interested in estimating the probability of occurring a death having the above-mentioned conditions. It is easy to verify that the ML estimate of this desired probability is:

$$\delta^{\mathrm{ML}}_{71,375} = P\left(X_7 = x_7^{(1)} | \pi_7 = \pi_7^{(375)}\right) = P\left(X_7 = 1 | \pi_7 = (3, 2, 6, 3)\right) = 0.0012$$

Therefore, the chance of death while wearing seat belt and driving a car, knowing that the driver has a diploma degree and a type 2 driving license, is 0.12%. Comparing the chance of having an injury or death provided the mentioned knowledge, we observe that the chance of having a death is 0.0154 less than the chance of being injured. In fact, it is at least 13 times of the chance of being injured. Obviously, the chance of being not injured for the assumed person would be $\delta^{\mathrm{ML}}_{73,375} = 1 - \delta^{\mathrm{ML}}_{71,375} - \delta^{\mathrm{ML}}_{72,375} = 0.9822$. Comparing this chance with the first probability, we observe that the chance of not being injured is 0.9656 more than that of being injured. In other words, the chance of not being injured is at least 59 times of the chance of being injured. It is also worth



7 Conditional probabilities of fatalities while not using seat belt in the road accident network

noting that the chance of not being injured is 0.9810 more than that of having a death. In fact, it is 818 times of the chance of a death. Figure 6 provides these conditional probabilities.

In addition to the computed probabilities, it is interesting to realize that the road accident BN can be used to examine the effect of changing status of some nodes to some alternatives. To make this clearer, once again, consider the situation that the car driver has a diploma degree and a type 2 driving license. To see the effect of using seat belt on reducing accident fatalities, one may compute the chance of being injured in an accident when assuming the driver does not wear her/his seat belt. Then, comparing the resulting probability with those computed under the assumption that the driver uses her/his seat belt, will demonstrate the effect of using seat belts in reducing accident fatalities.

Now, to compute the chance of being injured while not wearing seat belt for the mentioned driver, we observe that the special parent set would be $\pi_7^{(375)} = (3, 2, 6, 1)$ and the desired probability is $\theta_{72,125}$ which is estimated by its ML estimate as below:

$$\delta^{\mathrm{ML}}_{72,125} = P\left(X_7 = x_7^{(2)} | \pi_7 = \pi_7^{(125)}\right)$$
$$= P\left(X_7 = 2 | \pi_7 = (3, 2, 6, 1)\right) = 0.1140.$$

Comparing $\delta^{\mathrm{ML}}_{72,125}$ and $\delta^{\mathrm{ML}}_{72,375}$, reveals that the chance of being injured while wearing seat belt, i.e., 0.0166, can be increased to 0.1140, if the driver does not wear a seatbelt. In fact, the corresponding chance difference is 0.0974.

The chance of having a death while not wearing seat belt for the mentioned driver, i.e., $\theta_{71,125}$, is estimated by its ML estimate as provided below

$$\delta^{\mathrm{ML}}_{71,125} = P\left(X_7 = x_7^{(1)} | \pi_7 = \pi_7^{(125)}\right)$$
$$= P\left(X_7 = 1 | \pi_7 = (3, 2, 6, 1)\right) = 0.01$$

Comparing this value with $\delta^{\mathrm{ML}}_{71,375} = 0.0012$ reveals that the use of seat belt can decrease the death chance from 0.01 to 0.0012, which is 0.0088 in difference. Also, we observe that the chance of being not injured for the assumed person would be $\delta^{\mathrm{ML}}_{71,125} = 1 - \delta^{\mathrm{ML}}_{72,125} - \delta^{\mathrm{ML}}_{73,125} = 0.876$which is less than the change of being not injured while wearing a seat belt, i.e., 0.9822. Figure 7 provides the discussed conditional probabilities.

Taking the above chances in mind, one will confirm that the use of seat belts reduces number of accident fatalities significantly. It is interesting to realize that, in the learnt road accident network, there are 9473 conditional probabilities which can be computed similar to the above probabilities. Clearly, each one of the probabilities, will have its explanation and can be computed if needed.

Conclusion

The BN technique is an effective tool in analyzing the factors influencing the severity of road accident. In this paper, we considered learning the structure and parameters of the road accident data. For structure learning, we focused on the constraint-based approach that basically checks the data for conditional independence relationships. We implemented the PC algorithm to express the complicated relationship between the traffic accident and the causes. The results of analysis provide the valuable information on how to reveal the traffic accident causality. Although there are some interesting probabilities, which one might compute to make a decision in line with accident fatality reduction, the network persuades us to compute probability of being injured for a driver given some information about its node. We also investigated the effect of using seat belt on the number of accident fatalities. The results show that the use of seat belt can significantly decrease the number of accident fatalities. Finally, we emphasize that, if someone is interested in some other probabilities, as the network structure is at hand, they can easily compute the required probabilities to make inferences upon their choices.

Acknowledgements

The authors are grateful to Research Center of Tehran Traffic Police for providing the accident data. The authors are also grateful to the editor and anonymous reviewers for several helpful suggestions which significantly improved the quality of this paper.

References

Buntine, W. L. 1991. Theory refinement on Bayesian networks, in Proceedings of the 7-th International Conference Uncertainty in Artificial Intelligence, (eds. B. D, Ambrosio, P. Smets and P. Bonissone), 52–60, San Francisco, CA, Morgan Kaufmann Publishers.

Cano, A., Gómez-Olmedo, M. and Moral, S. 2008. A score based ranking of the edges for the PC algorithm, in Proceedings of the Fourth European Workshop on Probabilistic Graphical Models, 41–48, Hirtshals.

Cheng, J., Greiner, R., Kelly, J., Bell, D. and Liu, W. 2002. Learning Bayesian networks from data: an information-theory based approach, *Artif. Intell.*, **137**, 43–90.

Cooper, G. F. and Herskovits, E. H. 1991. *A Bayesian method for the induction of probabilistic networks from data*, Pittsburgh, PA, University of Pittsburgh, Section of Medical Informatics.

Davis, G. A. and Pei, J. 2003. Bayesian networks and traffic accident reconstruction, in Proceedings of the 9th International Conference on Artificial Intelligence and Law, 171–176, New York, NY, ACM.

de Campos, C. and Qiang, J. 2008. Improving Bayesian network parameter learning using constraints, in International Conference in Pattern Recognition, Tampa, FL.

de Morais, S. R. 2009. Bayesian network structure learning with applications in feature selection (PhD thesis), University of Lyon, Lyon.

de Oña, J., Mujalli, R. O. and Calvo, F. J. 2011. Analysis of traffic accident injury severity on Spanish rural highways using Bayesian networks, *Accid. Anal. Prev.*, **43**, (1), 402–411.

Deublein, M., Matthias, S. and Bryan, T. A. 2014. Prediction of road accidents: comparison of two Bayesian methods, *Struct. Infrastruct. Eng.*, **10**, (11), 1394–1416.

Eggermont, P. P. B. and LaRiccia, V. N. 2001. Maximum penalized likelihood estimation: volume I: density estimation, New York, Springer Science+Business Media.

Gregoriades, A. 2007. Towards a user-centred road safety management method based on road traffic simulation, in Simulation Conference, 2007 Winter, 1905–1914, Washington, DC, IEEE.

Heckerman, D., Geiger, D. and Chickering, D. M. 1995. Learning Bayesian networks: the combination of knowledge and statistical data (Technical Report MSR-TR-94-09), Redmond, Microsoft Research.

Heckerman, D., Meek, C. and Cooper, G. 1999. A Bayesian approach to causal discovery, in *Computation, causation, and discovery*, (eds. C. Glymour and G. F. Cooper), Menlo Park, CA, AAAI Press.

Hongguo, X., Huiyong, Z. and Fang, Z. 2010. Bayesian network-based road traffic accident causality analysis, in Information Engineering (ICIE), 2010 WASE International Conference, Vol. 3, 413–417, Beidaihe, IEEE.

Janssens, D., Wets, G., Brijs, T., Vanhoof, K., Arentze, T. and Timmermans, H. 2006. Integrating Bayesian networks and decision trees in a sequential rule-based transportation model, *Eur. J. Oper. Res.*, **175**, (1), 16–34.

Jensen, F. V. and Nielsen, T. D. 2007. Bayesian networks and decision graphs, New York, NY, Springer.

Karimnezhad, A. and Moradi, F. 2015. Bayesian parameter learning with an application, Accepted to Appear in *Metron*. doi: 10.1007/s40300-015-0077-0.

Koski, T. and Noble, J. M. 2009. Bayesian networks – an introduction, Chichester, Wiley.

Marsh, W. and Bearfield, G. 2004. Using Bayesian networks to model accident causation in the UK railway industry, in *Probabilistic safety assessment and management*, (eds. C. Spitzer *et al.*), 3597–3602, London, Springer.

Millar, R. B. 2011. Maximum likelihood estimation and inference, Chichester, Wiley.

Ozbay, K. and Noyan, N. 2006. Estimation of incident clearance times using Bayesian NETWORKS approach, *Accid. Anal. Prev.*, **38**, (3), 542–555.

Pearl, J. 1988. Probabilistic reasoning in intelligent systems, San Mateo, CA: Morgan Kaufmann.

Ramoni, M. and Sebastiani, P. 2003. Bayesian methods, in *Intelligent data analysis. An introduction*, (eds. M. Berthold and D. Hand), 131–168, New York, NY, Springer.

Simoncic, M. 2004. A Bayesian network model of two-car accidents, *J. Trans. Stat.*, **7**, (2–3), 13–25.

Spirtes, P., Glymour, C., and Scheines, R. 2000. Causation, prediction, and search (2nd ed.), Cambridge, MA, MIT Press.

Wach, W. 2013. Structural reliability of road accidents reconstruction, *Forensic Sci. Int.*, **228**, (1–3), 83–93.

Xu, Z. X., Jiang, Y., Lin, F. and Dai, L. 2011. The analysis and prevent in traffic accidents based on Bayesian network, *Adv. Eng. Forum*, **1**, 21–25.