

Oscar Petersson, Johan Levinsson

## **TDDC17 Lab3**

Högskoleingenjörsutbildning i datateknik, 180 hp

Laboration report - September 12, 2016

**Artificial Intelligence**

TDDC17, Linköpings universitet

Assistant:  
Josef Fagerström

IDA

## II : 5

a) What is the risk of melt-down in the power plant during a day if no observations have been made?

$$P(\text{Meltdown}) = 0.02578$$

What if there is icy weather?

$$P(\text{Meltdown}|\text{IcyWeather}) = 0.03472$$

b) Suppose that both warning sensors indicate failure. What is the risk of a meltdown in that case? Compare this result with the risk of a melt-down when there is an actual pump failure and water leak. What is the difference? The answers must be expressed as conditional probabilities of the observed variables,  $P(\text{Meltdown}|\dots)$ .

$$P(\text{Meltdown}|\text{PumpFailureWarning}, \text{WaterLeakWarning}) = 0.14535$$

$$P(\text{Meltdown}|\text{PumpFailure}, \text{WaterLeak}) = 0.2$$

$$P(\text{Meltdown}|\text{PumpFailure} \wedge \text{WaterLeak}) = 0.2$$

c) The conditional probabilities for the stochastic variables are often estimated by repeated experiments or observations. Why is it sometimes very difficult to get accurate numbers for these? What conditional probabilities in the model of the plant do you think are difficult or impossible to estimate?

Estimations are dependent on large datasets. The warnings could possibly be triggered by failure emulation, and could be fairly easy to get reliable estimates for. The weather and water leaks could be harder depending on where the power station is located, and pump failures (given that they are uncommon) would be hard to get reliable estimates for, as well as meltdowns as they are (hopefully, and usually) very rare.

d) Assume that the "IcyWeather" variable is changed to a more accurate "Temperature" variable instead (don't change your model). What are the different alternatives for the domain of this variable? What will happen with the probability distribution of  $P(\text{WaterLeak}|\text{Temperature})$  in each alternative?

A "generic"  $P(\text{Temperature} < 0)$  would give us a rough estimation, but more and finer ranges in common ranges would give an increased possibility for more exact estimations since a temperature of -20 most likely means that it has been cold for a longer time and will remain so for a while more, increasing the risk of waterleaks compared to a temperature of -1. This would let the probability distribution span between several discrete values rather than the current  $\{.1, .2\}$ .

## II : 6

a) *What does a probability table in a Bayesian network represent?* The probability table represent the probability of possible states in a node depending on its dependencies' states if such exists. If no dependencies, it shows the probability of possible states depending on randomness or unknown factors (such as IcyWeather: we don't know/care what actually causes—well, sub-zero temperatures, but ...— it, we just care that it is a 5% chance of IcyWeather).

b) *What is a joint probability distribution? Using the chain rule on the structure of the Bayesian network to rewrite the joint distribution as a product of  $P(\text{child} \mid \text{parent})$  expressions, calculate manually the particular entry in the joint distribution of  $P(\text{Meltdown}=F, \text{PumpFailureWarning}=F, \text{PumpFailure}=F, \text{WaterLeakWarning}=F, \text{WaterLeak}=F, \text{IcyWeather}=F)$ . Is this a common state for the nuclear plant to be in?*

A joint probability distribution gives the probability of each state for given variables. For  $n$  variables, each entry in an  $n$ -dimensional matrix represent one atomic state.

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i | X_{i-1}, \dots, X_1)$$

$$\begin{aligned} & P(\neg M, \neg PFW, \neg PF, \neg WLW, \neg WL, \neg IW) \\ &= P(\neg M | \neg PFW, \neg PF, \neg WLW, \neg WL, \neg IW) \cdot \\ & \quad P(\neg PFW | \neg PF, \neg WLW, \neg WL, \neg IW) \cdot P(\neg PF | \neg WLW, \neg WL, \neg IW) \cdot \\ & \quad P(\neg WLW | \neg WL, \neg IW) \cdot P(\neg WL | \neg IW) \cdot P(\neg IW) \\ &= P(\neg M | \neg PF, \neg WL) \cdot P(\neg PFW | \neg PF) \cdot P(\neg PF) P(\neg WLW | \neg WL) \cdot \\ & \quad P(\neg WL | \neg IW) P(\neg IW) \\ &= 0.999 \cdot 0.95 \cdot 0.9 \cdot 0.95 \cdot 0.9 \cdot 0.95 = 0.69378 \end{aligned}$$

c) *What is the probability of a meltdown if you know that there is both a water leak and a pump failure? Would knowing the state of any other variable matter? Explain your reasoning!*

The probability of  $P(\text{Meltdown} | \text{PumpFailure}, \text{WaterLeak})$  is 0.2 according to the probability table. *Meltdown* is only conditionally dependent on *PumpFailure* and *WaterLeak*, i.e. not dependent on any other variable. Hence, states of any other variable does not matter, which is as it should ... A warning light does not cause a meltdown, but the cause of the warning light might, and waterleaks could occur due to other reasons than frozen pipes.

d) *Calculate manually the probability of a meltdown when you happen to know that PumpFailureWarning=F, WaterLeak=F, WaterLeakWarning=F and IcyWeather=F but you are not really sure about a pump failure.*

Inference by Enumeration

$$P(X|e) = \alpha \cdot P(X, e) = \alpha \cdot \sum_y P(X, e, y)$$

$X = \{Meltdown\}$

$E = \{PumpFailureWarning, WaterLeak, WaterLeakWarning, IcyWeater\}$

$e = \{\neg PFW, \neg WL, \neg WLW, \neg IW\}$

$Y = \{PumpFailure\}$

$$\begin{aligned} P(X|e) &= \alpha \cdot P(X, e) = \alpha \cdot \sum_y P(X, e, y) \\ &= \alpha \cdot \sum_{PF} (P(M|PF = pf, \neg wl) \cdot P(\neg pfw|PF = pf) \cdot P(PF = pf) \cdot P(\neg wlw|\neg wl) \cdot P(\neg wl|\neg iw) \cdot P(\neg iw)) \\ &= \alpha \cdot \underbrace{P(\neg wlw|\neg wl) \cdot P(\neg wl|\neg iw) \cdot P(\neg iw)}_{\text{const}} \cdot \sum_{pf} (P(M|PF = pf, \neg wl) \cdot P(\neg pfw|PF = pf) \cdot P(PF = pf)) \\ &= \alpha \cdot (P(M|pf, \neg wl) \cdot P(\neg pfw|pf) \cdot P(pf) + P(M|\neg pf, \neg wl) \cdot P(\neg pfw|\neg pf) \cdot P(\neg pf)) \\ &= \alpha \cdot ((\langle 0.15, 0.85 \rangle \cdot 0.1 \cdot 0.1) + (\langle 0.001, 0.999 \rangle \cdot 0.95 \cdot 0.9)) = \alpha \cdot \langle 0.00236, 0.86265 \rangle \\ &= \frac{\langle 0.00236, 0.86265 \rangle}{0.00236 + 0.86265} = \langle 0.00272, 0.99728 \rangle \end{aligned}$$

### III : 1

Model the car with the applet tool and integrate it with the model of the plant.

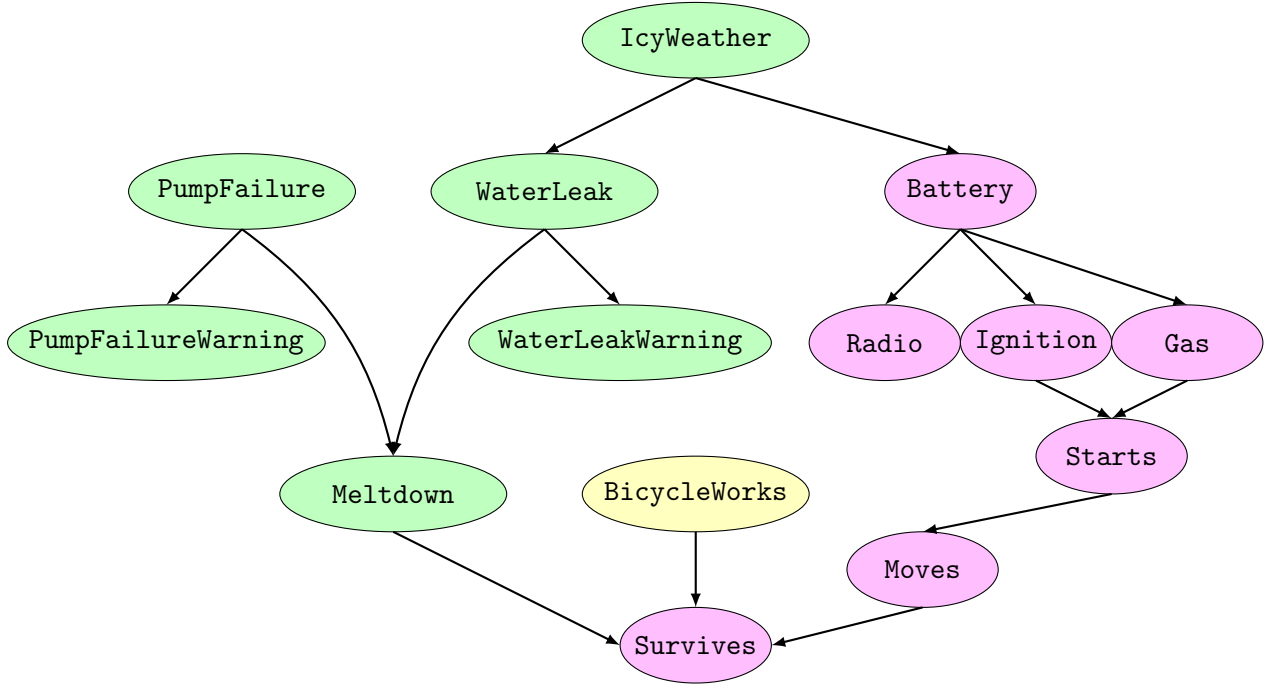


Figure 1: Bayesian network for survival probability in regards of vehicle and powerplant status.

a) During the lunch break, the owner tries to show off for his employees by demonstrating the many features of his car stereo. To everyone's disappointment, it doesn't work. How did the owner's chances of surviving the day change after this observation?

A radio that does not work is bad mojo, it might mean that the battery is out of juice which would reduce the probability of survival. Somewhat counter-intuitive to the commons, it decreases the probability from 0.99001 to 0.98116. Bayes' Theorem gives us:

$$P(\text{Battery}|\neg\text{Radio}) = \frac{P(\neg\text{Radio}|\text{Battery})P(\text{Battery})}{P(\neg\text{Radio})}$$

b) The owner buys a new bicycle that he brings to work every day. The bicycle has the following properties:

- $P(\text{BicycleWorks}) = 0.9$
- $P(\text{Survives}|\neg\text{Moves} \wedge \text{Meltdown} \wedge \text{BicycleWorks}) = 0.6$
- $P(\text{Survives}|\text{Moves} \wedge \text{Meltdown} \wedge \text{BicycleWorks}) = 0.9$

How does the bicycle change the owner's chances of survival?

The bicycle adds a component of redundancy to the possibilities of escape. Adding up the numbers, the survivability from 0.99001 to 0.99505.

c) It is possible to model any function in propositional logic with Bayesian Networks. What does this fact say about the complexity of exact inference in Bayesian Networks? What alternatives are there to exact inference?

In *singly connected* Bayesian networks, the time and space complexity is linear to the size of the network. In *multiply connected* networks, complexity is exponential. Alternatives to exact inference are *approximate inference* methods. Using randomized samples, the accuracy is depending on the amount of generated samples.