Fundamentals of AI and KR

Module 3: probabilistic and uncertain reasoning

Paolo Torroni

January 17, 2025 [70 minutes]

Customer churn refers to the phenomenon where customers stop doing business with a company or cancel their subscriptions to its products or services. It is a major problem for companies with many customers, like telecommunication providers. Some experts modeled customer churn in that sector in the following way:

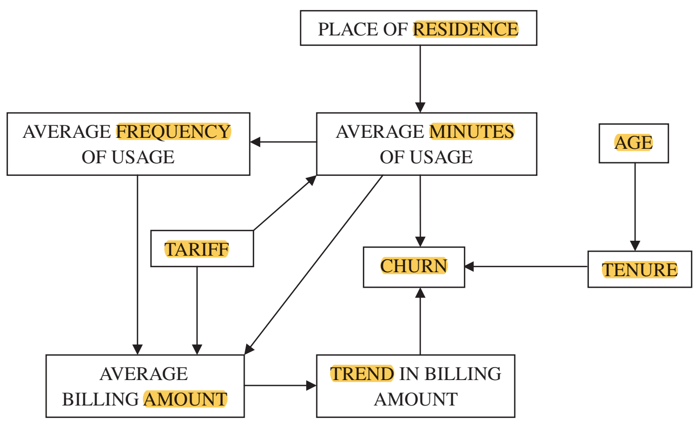

Figure 1: Bayesian Network defined by Pınar Kişioglu and Y. İlker Topcu*

The variables, all discretized, are:

- *Residence* $\in \{rural, urban\}$
- *Frequency* $\in \{low, middle, high\}$
- *Minutes* $\in \{short, medium, long\}$
- *Age* $\in \{young, old\}$
- *Tariff* $\in \{t0, t1, t2\}$ (three types of tariffs, for companies or individuals)
- *Churn* $\in \{no, yes\}$
- *Tenure* $\in \{short, medium, long\}$ (for how long one has been a customer)
- *Amount* $\in \{low, middle, high\}$
- *Trend* $\in \{downward, constant, upward\}$

* Applying Bayesian Belief Network approach to customer churn analysis: A case study on the telecom industry of Turkey,ESWA 2021

By analyzing their data, the experts observed that the conditional probability of *Churn* has the highest value when the average *Minutes* of usage is low, the *Tenure* is short, and the *Trend* in billing amount is downward.

Questions

1. *[1 pt]* Is the network a polytree?

2. *[1 pt]* Why do the experts relate the probability of *Churn* with *Minutes*, *Tenure* and *Trend*, and do not consider at the same time any other variables, like for example *Tariff* or *Age*?

3. *[1 pt]* How many independent parameters are needed to define the CPT for *Amount*?

4. *[3.5 pt]* Use **variable elimination** to calculate the probability of customer *Churn* assuming *short* *Tenure*, *low* average billing *Amount*, and *short* average *Minutes* of usage. Define the relevant CPT values as you consider appropriate.

5. *[1.5 pt]* Assume you would estimate the same probability distribution as above, by **Gibbs sampling**. Illustrate how to draw a sample, starting from a state of your choice. Base your answer on the same CPTs you defined to answer the previous point, and on the following random sequence: .11, .93, .28, .53, .05, . . .

6. *[3 pt]* Explain **Noisy-OR**: what it is, when it is applicable, how it helps.

Answers (including some extra comments)

1. Not it isn't a polytree, since there are multiple paths connecting *Tariff* and *Churn* (for example).

2. *Minutes*, *Tenure* and *Trend* happen to be *Churn*'s Markov blanket, hence *Churn*'s probability distribution is completely described by them. Moreover, by local semantics, the probability distribution of a node is described as a conditional distribution given the parents only.

3. *Amount* is a discrete variable with 3 possible values and 3 parents, all with arity 3, hence: $(3-1) \times 3^3 =$ **54** independent parameters.

4. **P**(*C*|*M*=*short*, *Te*=*short*, *A*=*low*)'s **ancestral tree** does not help eliminate nodes. However, by looking at independencies, we find that *Residence*, *Frequency* and *Tariff* can be ignored. Hence,

$$\mathbf{P}(C|M{=}short,\ Te{=}short,\ A{=}low) =$$
$$\alpha\,\mathbf{P}(C|M{=}short,\ Te{=}short,\ Tr) \times \mathbf{P}(Tr|A{=}low) =$$
$$\alpha \sum_{tr\in\{down,const,up\}} \mathbf{P}(C|M{=}short,\ Te{=}short,\ tr) \times \mathbf{P}(tr|A{=}low)$$

A possible CPT for **P**(*C*|*M*=*short*, *Te*=*short*, *Tr*) is:

	P(*C*=*no*|*M*=*sh*,*Te*=*sh*,*Tr*)	**P**(*C*=*yes*|*M*=*sh*,*Te*=*sh*,*Tr*)
Tr=*down*	0.4 (x)	0.6 (1-x)
Tr=*const*	0.6 (y)	0.4 (1-y)
Tr=*up*	0.9 (z)	0.1 (1-z)

A possible CPT for **P**(*Tr*|*Am*=*low*) is:

	P(*Tr*|*Am*=*low*)
Tr=*down*	0.4 (v)
Tr=*const*	0.5 (w)
Tr=*up*	0.1 (1-v-w)

With these values, we obtain:

$$f_1(C) = \sum_{tr\in\{d,c,u\}} \mathbf{P}(C|M{=}sh,Te{=}sh,,tr)\ \mathbf{P}(Tr|A{=}low) =$$

	$f_1(C)$
C=*no*	x·v+y·w+z·(1-v-w) = 0.16 + 0.3 + 0.09 = 0.55
C=*yes*	((1-x)·v+(1-y)·w+(1-z)·(1-v-w) = 0.24 + 0.2 + 0.01 = 0.45

Hence, **P**(*C*|*M*=*short*, *Te*=*short*, *A*=*low*) $= \langle 0.55, 0.45\rangle$

5. Let us assume we start from a state where *M*=short, *Te*=short, and *A*=low, as per given evidence, and then *Tr*=down and *C*=yes. The other variables do not contribute to the probability distribution we have to compute, so we

can ignore them. We must first select a variable to sample. Say we select *C*. Its Markov blanket is {*M*, *Te*, *Tr*}. According to the table defined previously, we have that P(*C=no*|*M=short*, *Te=short*, *Tr=down*) = x = 0.4. From the sequence .11, .93, ... we need only one number: let's take the first one. Since 0.11 < x, we obtain a sample with *Tr*=down and *C*=no.

If we selected *Tr* instead of *C*, we would consider *mb(Tr)* = {*A*, *C*, *M*, *Te*}, and **P**(*Tr*|*A=low*, *C=yes*, *M=short*, *Te=short*) =
α **P**(*Tr*|*A=low*) **P**(*C=yes*|*M=short*, *Te=short*, *Tr*) =
α ⟨*v* × *(1-x)*, *w* × *(1-y)*, *(1-v-w)* × *(1-z)*⟩ =
$\alpha \langle 0.4 \times 0.6, 0.5 \times 0.4, 0.1 \times 0.1 \rangle =$
$\alpha \langle .24, .2, .01 \rangle = \langle 0.53, 0.44, 0.02 \rangle.$

Since 0.11 < 0.53, we would obtain a sample with *Tr*=down and *C*=yes.

6. Important points to cover:

   - *What*: a compact representation of a CPT. Probabilities are computed as 1 - failure probability of joint causes.
   - *When*: all the following conditions must be verified: (1) we have a common effect node and all the nodes (causes and effect) are binary; (2) all possible causes are represented, possibly with the addition of a leak node; (3) causes have independent failure probability.
   - *Why*: it reduces the number of parameters from exponential in the number of nodes to linear