Fundamentals of AI and KR
Module 3: probabilistic and uncertain reasoning

Paolo Torroni

July 7, 2020 [25 minutes for solving + 10 for uploading]

You are designing a new e-commerce system, to sell products online. You can display advertisements ("ads") of three types: book ads, toy ads, and holiday ads. However, your web site can only display one ad at a time. Your system should guess which type of ad has a higher chance of being clicked.

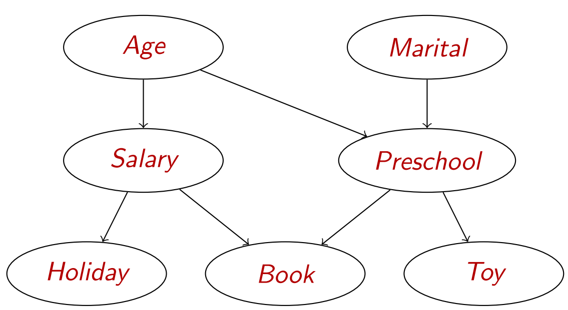
You don't have verified profile information on each customer. You can, however, guess at least some profile features, based on your domain knowledge and on your observations about which ads the customer does or doesn't click on ("clicking behaviour"). In particular, you have identified the following **profile features**:

- *Age*: the customer's age group (*a1*: <25, *a2*: $[25..45]$, *a3*: $46+$);
- *Marital*: the customer's marital status (*s*=single, *c*=couple);
- *Salary*: the customer's salary range (*low*, *mid*, *high*);
- *Preschool*: whether the customer has pre-school children (*yes*, *no*).

The customer's **clicking behaviour** is instead determined as follows:

- *Holiday*: the customer clicks on a holiday ad (h, $\neg h$);
- *Book*: the customer clicks on a book ad (b, $\neg b$);
- *Toy*: the customer clicks on a toy ad (t, $\neg t$).

In the end, you produce the following network:



Questions

1. What are the dimensions of the *Preschool* CPT?

2. You first display a sequence of ads and record which types of ads get clicked; then, you use that data to infer some of the customer's profile features. Which type of reasoning pattern are you following? Show a query that uses the customer's observed clicking behaviour as evidence.

3. You display a holiday ad, and observe that it doesn't get clicked. Can this observation help predicting whether the customer would click on a toy ad?

4. You expand your business to sell *Skateboard*s: an article that's especially popular with male customers in the *a1* age group. How would you expand your Bayesian network to include also skateboard ads?

5. You display a skateboard ad, and observe that your customer doesn't click on it. Can this observation alone help predict your customer's marital status?

6. The same customer joins a fidelity programme and fills in a form where she declares to be a female parent of a pre-school child. She chooses not to disclose any other information about herself. Given this new evidence, *does her not clicking on a skateboard ad give you any information about her marital status*?

7. With reference to the initial network without skateboard, consider the following query: P(b|h,no). Choose an appropriate sampling method and show how to generate a sample for such a method.

   Notice: you don't have to generate a full sample: just show how you would generate it (some steps that illustrate the procedure are enough). You can make any assumptions you consider relevant and appropriate about CPT values, state of the network, randomly generated numbers, etc.

Answers (including some extra comments)

1. *Preschool* is Boolean and has 2 parents with 3 and 2 possible values, therefore its size is $3 \times 2 \times 2$ (6 independent and 6 dependent parameters).

2. We use evidential reasoning: our aim is to obtain probability distributions for *Age*, *Marital*, *Salary* and *Preschool* that explain our observations about ad clicks.

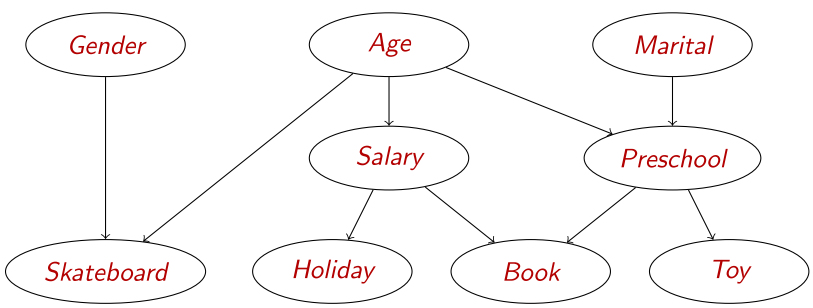
   For example, if we show a holiday ad and a toy ad to a customer, and the customer clicks on the toy ad but not on the holiday ad, we can infer the distributions of the customer's salary and parental status as a conjunctive query: P(S,P|¬h,t).

3. Yes, in general this observation may help, because there's an active trail between *Holiday* and *Toy*, therefore P(T|¬h)≠P(T).

   Of course that's true under normal conditions (assuming non zero probabilities in the CPTs), and as long as salary and parental status are unknown, and either age is unknown or book is known.

   When we use the network to predict the customer's clicking behaviour, we are doing causal reasoning.

4. Here is one possible way of extending the network:



5. With no further evidence, there's no active trail between *Skateboard* and *Marital*, therefore that observation alone doesn't help.

6. Again, under normal conditions, the trail *Skateboard* ← *Age* → *Preschool* ← *Marital* is now active, so the short answer is: yes, in general it does.

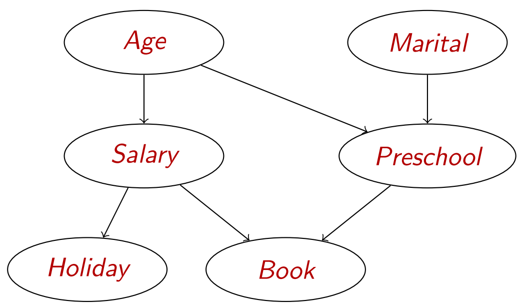
   We may also want to consider the actual CPTs for *Skateboard*. For example, if skateboards were *only* popular among male customers, so much so that the probability that a female customer clicks on a skateboard ad is close to 0, then the fact that a *female* customer does not click on a skateboard ad may bring no information about her marital status.

   For example, if the CPT for P(sk|G,A) is:

m	$a1$	0.10
m	$a2$	0.01
m	$a3$	0.02
f	$a1$	0.00
f	$a2$	0.00
f	$a3$	0.00

then, P(*M*|*f,yes,¬sk*)=P(*M*|*f,yes*). So, another possible answer is: it may, but it really depends on the CPTs.

7. First of all, let's notice that *Book* is independent of *Toy* given *Preschool*. Therefore, we can consider a simplified subgraph without *Toy*:



One first option is to use rejection sampling. Let's sample our variables in the following order: *Age*, *Salary*, *Marital*, *Preschool*, *Holiday*, *Book*. Let's assume that we generate the following sequence of random numbers:

0.38, 0.92, 0.1, 0.02, 0.49, 0.19, ...

We take the first random number. Assume that P(A) = ⟨ 1/3, 1/3, 1/3 ⟩: then, 0.38 yields *Age=a2*. Assume that P(S|a2) = ⟨ 1/6, 1/2, 1/3 ⟩: 0.92 yields *Salary=high*. Assume that P(M) = ⟨ 0.3, 0.7 ⟩: 0.1 yields *Marital=s*. Next, we consider *Preschool*, which is an evidence variable. Assume that P(P|a2,s) = ⟨ 0.15, 0.85 ⟩: 0.02 yields *yes*. This disagrees with the evidence, so we reject the sample and use the next random number, 0.49, to start generating a new sample.

If we had opted for likelihood weighting, instead of rejecting the sample, we would have proceeded to sample *Book* using 0.02, and then weighted the sample by P(*no*|a2,s) P(*h*|high) = 0.85 × . . . .

Finally, since our evidence is rather down in the tree, instead of rejection sampling or likelihood weighting we may want to use Gibbs sampling. Let's illustrate. Assume that we start from the following random state:

$$\langle A, S, M, \underline{P}, \underline{H}, B \rangle = \langle a1, mid, s, \underline{no}, \underline{h}, \neg b \rangle.$$

We arbitrarily pick a variable that is not an evidence variable, say *S*. Its Markov blanket is $\{A,\ \underline{P},\ \underline{H},\ B\ \}$, with values *a1, no, h,* $\neg b$, so the next sample will be $\langle\ a1,\ X,\ s,\ \underline{no},\ \underline{h},\ \neg b\ \rangle$, where $X \in \{\ low,\ mid,\ high\ \}$ is determined by sampling from the distribution $\mathrm{P(S|mb(S)) = P(S|a1)\ P(h|S)\ P(\neg b|S,no)}$ using the first random number, 0.38.