Fundamentals of AI and KR

Module 3: probabilistic and uncertain reasoning

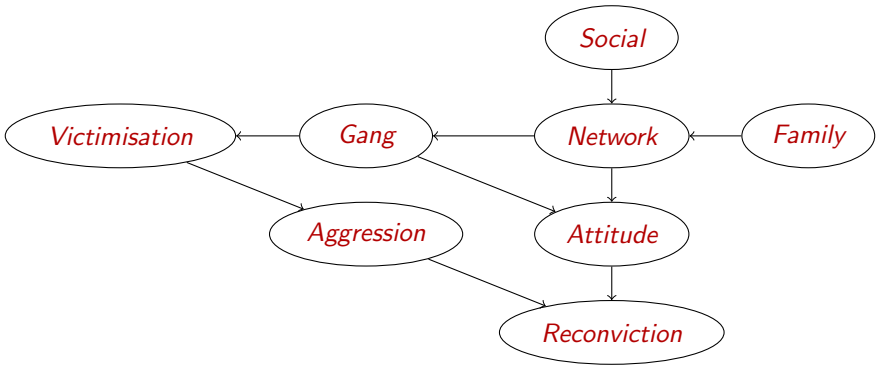
Paolo Torroni

January 26, 2020 [30 minutes for solving + 10 for uploading]

A group of forensic medics is building a Bayesian Network model to assess the risk of violent reoffending of convicted criminals, and decide possible interventions. They found out that involvement in criminal lifestyle is associated with violence, and is influenced by the presence of criminal activity in the family or peer groups, which in turn may foster attitudes supportive of crime.

Their convict model uses the following Boolean variables and connections:

- *Social*: presence of averse socio-economic factors
- *Family*: presence of a criminal family background
- *Network*: involvement in a criminal network
- *Gang*: membership in a criminal gang
- *Victimisation*: feeling of victimisation leading to violent thoughts
- *Attitude*: danger level of negative/criminal attitude
- *Aggression*: danger level of aggressive behaviour
- *Reconviction*: violent reoffence leading to reconviction



Questions

1. Define possible CPTs for *Gang* and *Attitude*. Assume it **impossible** to be a member of a *Gang* **without** being involved in a criminal *Network*.
2. According to the model, can *Attitude* influence *Aggression* of a gang member? *Hint: a "gang member" is an individual for which Gang=true.*
3. Experts define two possible interventions: *Psychiatric* therapy, aimed to reduce the feeling of *Victimisation*, and *Employment* policies, aimed to improve *Social* factors. Expand the network to include these possible interventions. *Hint: each "intervention" can be modelled as a new Boolean variable.*
4. Which interventions could reduce the risk of *Reconviction* of subjects known to be involved in a criminal *Network*?
5. Illustrate a method to compute the probability of *Victimisation* given (1) a dangerous level of negative/criminal attitude, (2) averse socio-economic factors, and (3) no psychiatric therapy (*Attitude*=true, *Social*=true, *Psychiatric*=false). Base your illustration on the CPTs you defined earlier, assuming $P(family)$=0.4, and the following CPTs:

$P(network \mid social, family)$	
s, f	0.8
$s, \neg f$	0.5
$\neg s, f$	0.6
$\neg s, \neg f$	0.1

$P(victimisation \mid psychiatric, gang)$	
p, g	0.4
$p, \neg g$	0.2
$\neg p, g$	0.6
$\neg p, \neg g$	0.3

Notice: only sketch the important steps - no need to perform all calculations!

Answers (including some extra comments)

1. *Gang* is a Boolean variable with a Boolean parent, therefore we have a table with one independent column and two rows. Because it is not possible to be a member of a *Gang* without being involved in a criminal network, $P(gang|\neg network)$ must be 0. A possibility:

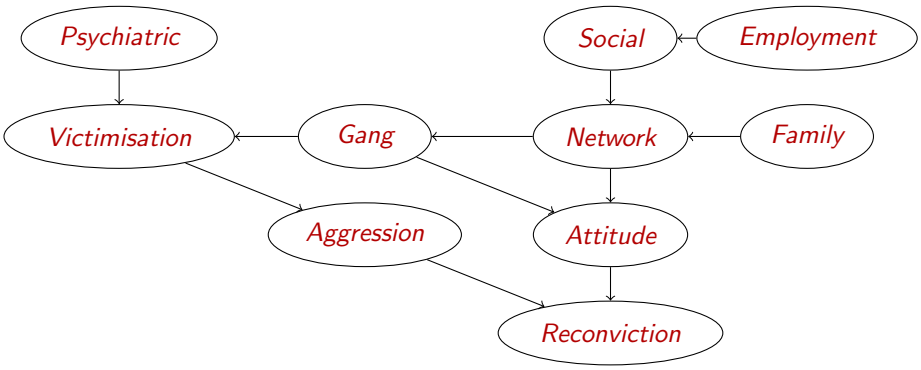
$P(gang\|network)$	
n	0.7
$\neg n$	0

   Attitude is a Boolean variable with two Boolean parents, therefore we have a table with one independent column and four rows. Assuming that *Network* and *Gang* positively contribute to *Attitude*, a possible CPT could be:

$P(attitude\|gang, network)$	
g, n	0.7
$g, \neg n$	undefined
$\neg g, n$	0.4
$\neg g, \neg n$	0.2

2. Since our only evidence is *Gang*=true, there's no active trail between *Attitude* and *Aggression*, so there can be no influence between them.

3. This is how the network should be expanded:



4. Our evidence, *Network*=true, blocks the flow of influence between *Employment* and *Reconviction*, therefore *Employment* policies wouldn't help.

   Psychiatric therapy instead may help, because there's an active trail from *Psychiatric* to *Reconviction* through *Victimisation* and *Aggression*.

5. First of all we should consider which variables contribute to the query. The set of query/evidence variables is {*Victimisation*, *Attitude*, *Social*, *Psychiatric*}. We can thus disregard *Aggression* and *Reconviction* as they do not belong to the ancestral tree. Moreover, *Victimisation* is independent of *Employment*

given *Network*, so we can disregard *Employment* as well. We are left with three relevant hidden variables: *Gang*, *Network*, and *Family*. Let's denote *Psychiatric*, *Victimisation*, *Gang*, *Attitude*, *Network*, *Social* and *Family* by *P*, *V*, *G*, *A*, *N*, *S* and *F* respectively. Our probability query can be expressed as:

$$P(V|a,s,\neg p) = \alpha \sum_g \sum_n \sum_f P(V|\neg p, g) P(g|n) P(a|g,n) P(n|s,f) P(f)$$

To evaluate that expression we can use variable elimination. Iteratively, we group all factors that depend on a given hidden variable, multiply them together, and sum out. We do that in turn for each relevant hidden variable. Let us (arbitrarily) choose the order of hidden variables: *F*, *N*, *G*.

$$\begin{aligned} P(V|a,s,\neg p) &= \alpha \sum_g P(V|\neg p, g) \sum_n P(g|n) P(a|g,n) \sum_f P(n|s,f) P(f) \\ &= \alpha \sum_g P(V|\neg p, g) \sum_n P(g|n) P(a|g,n) \underbrace{\sum_f f_N(n,f)}_{f_{\bar{F}N}(n)} \end{aligned}$$

We sum out *F*:

$f_{\bar{F}N}(n) = \langle 0.4 \times 0.8 + 0.6 \times 0.5, 0.4 \times 0.2 + 0.6 \times 0.5 \rangle = \langle 0.62, 0.38 \rangle$

$$P(V|a,s,\neg p) = \alpha \sum_g P(V|\neg p, g) \underbrace{\sum_n P(g|n) P(a|g,n) f_{\bar{F}N}(n)}_{f_{\bar{N}\bar{F}G}(g)}$$

We sum out *N*:

$f_{\bar{N}\bar{F}G}(g) = \langle 0.7 \times 0.7 \times 0.62, 0.3 \times 0.4 \times .62 + 1 \times 0, 2 \times 0.38 \rangle = \langle 0.3038, 0.1504 \rangle$

$$P(V|a,s,\neg p) = \underbrace{\alpha \sum_g P(V|\neg p, g) f_{\bar{N}\bar{F}G}(g)}_{f_{\bar{N}\bar{F}\bar{G}}(V)}$$

We sum out *G*:

$f_{\bar{N}\bar{F}\bar{G}}(V) = \langle 0.6 \times 0.3038 + 0.3 \times 0.1504, 0.4 \times 0.3038 + 0.7 \times 0.1504 \rangle = \langle 0.2274, 0.2268 \rangle$

Finally, we renormalize and obtain:

$P(V|a,s,\neg p) = \langle 0.501, 0.499 \rangle$

According to the model, a high danger level of criminal attitude and averse socio-economic factors give an even chance of *Victimisation*, without psychiatric therapy. It can be verified that with the introduction of psychiatric therapy the chance of *Victimisation* would drop down to approxmately $1/3$.