Belief Updating in Bayesian Networks

By Yan Kang

- Conditional Probability Table (CPT) Updating
- Bayesian Network reasoning with uncertain evidences

- Conditional Probability Table (CPT) Updating

Jeffrey's Rule

Given a joint probability distribution Q(X) and let Z ⊆ X\Y, for any y we have the updated distribution

$$Q'(x) = \frac{Q(z,y)}{Q(y)} R(y) = \begin{cases} Q(x)\frac{R(y)}{Q(y)} & \text{if } Q(y) \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

Definition 1. The *I-divergence* (also known as Kullback-Leibler distance and relative entropy) between *P*(*X*) and *Q*(*X*) is given by

$$I(P \| Q) = \begin{cases} \sum_{P(x)>0} P(x)\log\frac{P(x)}{Q(x)} & \textit{if} \quad P << Q \\ +\infty & \textit{otherwise} \end{cases}$$

Definition 2. *Q*(*x*) is said to be an *I-projection* of *P*(*x*) on a convex set of distributions **S** = {p∈P,p(y) = R(y)},if

$$I(Q \| P) = \min_{\tilde{Q} \in S} I(\tilde{Q} \| P)$$

the I-projection of *P*(*x*) on S can be calculated by

$$Q(x) = \begin{cases} P(x)\cdot\frac{R(y)}{P(y)} & \textit{if } P(y) \neq 0 \\ 0 & \textit{otherwise} \end{cases}$$

It has been proven that by this formula , Q(x) is the I-projection of p(x) on **S**

IPFP

Iterative proportional fitting procedure (IPFP) extends this idea to modify P(X) with multiple constraints by continuously projecting the distribution resulted from the previous iteration to next constraint . This procedure is formally defined as follows.

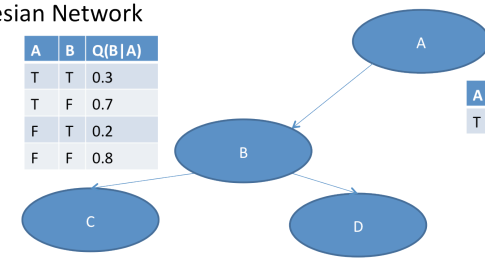
Definition 3. Let R={R(y^1),…,R(y^m)} be a set of constraints (or we can call it soft evidence) and Q_0(x) the initial distribution. Then for k=1,2,… , j = 1+(k-1) mod m and R(y^j) << Q_{k-1}(y^j)for all k, j, IPFP is defined by

$$Q_k(x) = \begin{cases} Q_{k-1}(x)\cdot\frac{R(y^j)}{Q_{k-1}(y^j)} & \textit{if} \quad Q_{k-1}(y^j) > 0 \\ 0 & \textit{otherwise} \end{cases}$$

m is the number of constraints, k is the iteration index, and j determines the constraint used at step k.

The IPFP has been proven to converge if there exists a distribution that satisfy all the constraints in R. and the converging distribution Q*(x) is the I-projection of initial distribution Q_0(x) on **S**.

Bayesian Network



A	B	Q(B|A)
T	T	0.3
T	F	0.7
F	T	0.2
F	F	0.8

A	Q(.)
T	0.2

A Bayesian Network N of n variables x = (x_1,x_2,…,x_n) have two parts:

(i) network structure Ns = {(x_i, π_i)}, which captures the interdependencies between variables.

(ii) a set of conditional probability table (CRTs) N_p = {Q(x_i | π_i)}, which represents the degree of interdependencies.

Definition. A joint distribution Q(x) is said to be structurally consistent with a Bayesian Network N if there exists a set of CPTs = {Q(xi|πi)} such that $Q(x) = \prod_{i=1}^{n} Q(x_i|\pi_i)$

Therefore, in order to let the converging distribution Q*(x) be consistent with the Bayesian Network, we treat the network structure as an additional constraint, which is:

$$R_{m+1}(x) = \Pi_{i=1}^{n} Q_{k-1}(x_i \mid \pi_i)$$

this version of IPFP is called Extended IPFP.

We can extract the Conditional Probability Tables from the converging distribution by any inference method. These CPTs are the ones that satisfy all the probability constraints that we have imposed on the Bayesian Network

- Bayesian Network Reasoning

Bayesian Network Reasoning is similar to CPTs updating.

The main difference:

CPTs updating: we are aim to update the CPTs of Bayesian Network.

Reasoning: we are aim to make inferences with uncertain evidences.

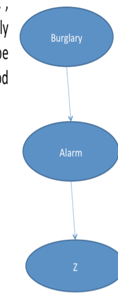
Jeffrey's rule is a natural choice for making inference with uncertain evidences *R*(*Y* ⊆ *X*). However, Jeffrey's rule cannot directly apply when the joint distribution is represented as a BN. This can be overcome by converting soft evidence to virtual evidence.

Consider a distribution Q(X) and a soft evidence (new evidence) R(Y), *Y* ⊆ *X*. All possible instantiations of Y, y_1, , …, ,y_n. ∈Y , (for example R(T=0.4, F=0.6)) form a mutually exclusive and exhaustive set of events. R(Y) then can be converted to a virtual evidence Z, with the likelihood ratios $\lambda_1 : \ldots : \lambda_n$.

$$Q(Z=T \mid Y=y_1) : Q(Z=T \mid Y=y_2) : \ldots : Q(Z=T \mid Y=y_n)$$
$$= \lambda_1 : \lambda_2 : \ldots : \lambda_n$$

$$\lambda_1 : \lambda_2 : \ldots : \lambda_n = \frac{R(y_1)}{Q(y_1)} : \frac{R(y_2)}{Q(y_2)} : \ldots : \frac{R(y_n)}{Q(y_n)}$$

A	Z	Q(Z|A)
True	True	λ_T = K*R(A=T)/Q(A=T) =K* 0.4/0.198
False	True	λ_F = K*R(A=F)/Q(A=F) =K* 0.6/0.802



Virtual Node

Burglary	Q(.)
True	0.2
False	0.8

Burglary	Alarm	Q(A|B)
True	True	0.95
True	False	0.05
False	True	0.01
False	False	0.99

A	Q(.)
T	0.198
F	0.802

A	R(.)
T	0.4
F	0.6

A	Z	Q(Z|A)
True	True	λ_T = K*R(A=T)/Q(A=T) =K* 0.4/0.198
False	True	λ_F = K*R(A=F)/Q(A=F) =K* 0.6/0.802

The new distribution obtained after accepting the soft evidence on y_1. ,…, ,y_n.

$$Q(ALARM \mid Z=T) = Q(x)\frac{\lambda_i}{\sum_j Q(y_j)\lambda_j} \quad \text{if x satisfy } y_j$$

we call this formula pearl's method. when substituting with its corresponding likelihood, the new distribution is exactly the same as the one obtained by applying R(Y) using the Jeffrey's rule

$$Q'(ALARM) = Q(ALARM \mid Z=T) = Q(ALARM)\frac{\frac{kR(y_i)}{Q(y_i)}}{\sum_j Q(y_j)\frac{kR(y_j)}{Q(y_j)}} = Q(ALARM)\frac{R(y_i)}{Q(y_i)}$$

$$Q(A=T \mid Z=T) = \frac{Q(A=T)\lambda_T}{Q(A=T)\lambda_T + Q(A=F)\lambda_F} = 0.4$$

$$Q(A=F \mid Z=T) = \frac{Q(A=F)\lambda_F}{Q(A=F)\lambda_F + Q(A=T)\lambda_T} = 0.6$$

So if we have N soft evidences, we can just add N virtual nodes to the Bayesian Network.

- Future work

The extended IPFP algorithm and BN-IPFP both require that all the constraints or soft evidences to be consistent with each other. We are now working on the algorithm that can deal with the situation that the constraints or soft evidences are inconsistent with each other.