Decision theory = utility theory (represent & infer preference) + prob theory

Maxmize expected utility: $a^* = \text{argmax}_a \sum_s P(s|a)U(s)$

Size of distribuT for n vars with domain size d : $d^n$

CSP: state whether assignM is possible

Marginal distribuT : eliminate Var MarginalizaT (sum out)

$X_1$的边缘分布 $P(X_1 = x_1) = \sum_{x_2} P(X_1 = x_1, X_2 = x_2)$

Conditional Prob: $\boxed{P(a|b) = P(a,b)/P(b)}$

Evidence var: $e_1 \cdots e_k$
Query var: Q
hidden var: $H_1 \cdots H_r$
→ All var: $X_1 \cdots X_n$

Inference by EnumeraT: (aim: $P(Q|e_1 \cdots e_k)$)

①Select entries (v evidence) ②sum out H得 $P(Q, e_1 \cdots e_k)$ ③normalize

worst case: $O(d^n)$ time; $O(d^n)$ space to store joint distri.

Chain rule: $P(a,b,c) = P(a)P(b|a)P(c|a,b)$.

$$P(x_1, x_2 \cdots x_n) = \prod_i P(x_i | x_1 \cdots x_{i-1})$$

Bayes' rule $P(x|y) = \frac{P(y|x)}{P(y)} P(x)$; Bayes NetWork: x, y indepdent if $\forall x,y$ $P(x,y) = P(x)P(y)$

Conditional Independence: $X \perp\!\!\!\perp Y | Z$ (x,y independent given z)

iff $\forall x,y,z$ $P(x|y,z) = P(x|z)$ or iff $\forall x,y,z$ $P(x,y|z) = P(x|z)P(y|z)$

BN: 有向无环、Conditional distri for each node give its parent Var in Graph

CPT: conditional probability table: each row is a distribution for child given a configuration of its parents

A Bayes net = Graph (Topo) + local Conditional Probs.

# free parameter = $(d-1)\prod_i d_i$ (each Table row sum → 1)

$d_i$: Parent domain Size. d: Child domain size

n var, max domain size = d, max #parents = k, full joint distri = $O(d^n)$

BN size = $O(n \cdot d^{k+1})$ $\boxed{P(X_1, X_2 \cdots X_n) = \prod_i P(X_i | Parents(X_i))}$

- given parents, var $\perp\!\!\!\perp$ non-descendants (conditional indep).

Markov-blanket 包括 Parents, Children, children's parents

- var $\perp\!\!\!\perp$ all other var given Markov-blanket

$X \to Y \to Z$ : X NOT$\perp\!\!\!\perp$ Z, $(X \perp\!\!\!\perp Z)|Y$ (Proof: $P(z|x,y) \overset{\text{chain rule}}{=} P(z|y)$)

$X \leftarrow Y \to Z$ : X NOT$\perp\!\!\!\perp$ Z, $(X \perp\!\!\!\perp Z)|Y$ $P(x,y,z) = P(y)P(x|y)P(z|y)$

$X \to Y \leftarrow Z$ : $\underbrace{X \perp\!\!\!\perp Z}_{\text{Active}}$, $\underbrace{(X \text{ NOT} \perp\!\!\!\perp Z)|Y}_{\text{blocked}}$ X is "d-seperated" from Y & Z if all path from X to Y are blocked

$(X \perp\!\!\!\perp Y)|Z$ if "d-seperated"

Markov network also encodes joint distri but 图无向

A factor is a multi-dimensional array to represent $P(Y_1 \cdots Y_N | X_1 \cdots X_M)$

①Joint distri $P(X,Y)$

②Selected joint $P(x,Y)$ fixed x, sums to $P(x)$

③Single conditional $P(Y|x)$ fixed x, sums to 1

④Family of conds $P(X|Y)$ all X, Y, sums to |Y|

⑤Specified family $P(y|X)$ sums to?

Join Factors: Ⓡ→Ⓣ $\forall r,t$: $P(r,t) = P(r) \cdot P(t|r)$


Eliminate: $P(R,T)$ → [Sum out (marginalize) R] → $P(T)$

i.e. join $P(R), P(T|R), P(L|T)$ 得 $P(R,T,L)$

Var EliminaT: Initial fac → while hidden var {Pick H; join all fac mentionG H; sum out H} → Join all remainG fac and norm $\frac{1}{Z}$

Diff: Inference by Enum: Join all then Eliminate one by one

Var Elim : Join on var1 → Elim. var1 → Join var2 → Elim V2

Var EliminaT OrderG: ①Z, $X_1 \cdots X_n$ : $2^{n+1}$ ②$X_1, X_2 \cdots X_n$, Z : $2^2$

X Exist an orderG that only results in small factors

3SAT: $(a_1 \vee b_1 \vee c_1) \wedge (a_2 \vee b_2 \vee c_2) \cdots \wedge (a_n \vee b_n \vee c_n)$, $P(+z) = \sum_{xy} P(x,y,+z)$ $= \sum_{x, s.t. z=T} P(x)$

$P(z) > 0$ iff Sentence is satisfiable → NP-hard

$P(z) = S \times 0.5^7$ where S = # satisfyG assignM, 7 = #x

Complexity of VE is linear in BN size (# CPT Entries)

Eliminate from the leaves to root

## Message passing on general G (Junction tree Algo)

Loopy Belief Propagation. Simply pass msg on general tree

X terminate with loops; run until convergence (x guarantee)

approximate but tractable on large G.

## Sampling

O(n) mem Prior Sampling: ①sample $X_i$ from $P(X_i | P_{ar}(X_i))$

$P(W) = \langle w: 0.8, \neg w: 0.2 \rangle$ (get counts than normalize)

$S_{PS}(x_1 \cdots x_n) = \prod_{i=1}^{n} P(x_i | Parents(X_i)) = P(x_1 \cdots x_n)$

Rejection Sampling: if $P(C|r,w)$ when get Sample, reject if not r = w = true

after①, if $x_i$ × consistent with evidence, reject and return

likelihood weightG: fix evidence, sample the rest to be consist, weight each sample by p of evidence var give parents

for i = 1..n { if $X_i$ is evidence var, { $x_i$ = observed val; for $X_i$, set $w = w * P(x_i | Par(i))$ } else { sample $x_i$ from $P(X_i | Par(X_i))$ } } return $(x_1 \cdots x_n)$, w

$S_{WS}(z,e) = \prod_{i=1}^{l} P(z_i | Par(Z_i))$ (z is sampled, e is fixed evidence.

$w(z,e) = \prod_{i=1}^{m} P(e_i | Par(E_i))$ ⇒ Consist: $S_{WS}(z,e) \cdot w(z,e) = P(z,e)$

Gibbs Sample: $X_i' \sim P(X_i | x_1, \cdots x_n) \overset{BN}{=} P(X_i | Markov\_blanket(X_i))$

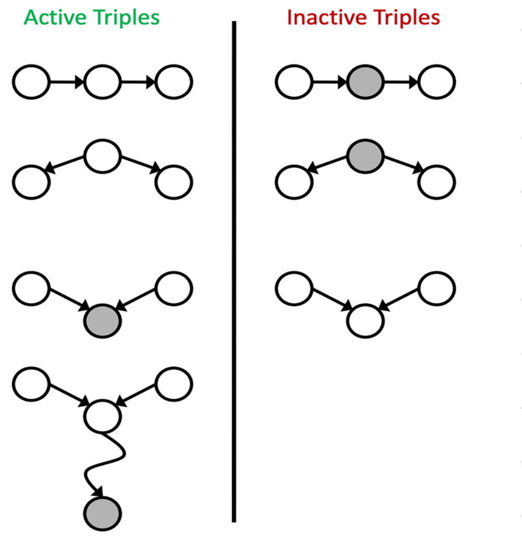
fix evidence → initial all → chose non-e var x

resample $X \sim P(X | Markov\_blanket(X))$

MCMC: Markov Chain Monte Carlo (sampling by ConstructG MChain)

Metropolis-Hastings: sample from $g(x'|x)$

accept with Prob = $\min(1, \frac{P(x')g(x|x')}{P(x)g(x'|x)})$ Prob = 1 in Gibbs S




| | |
|---|---|
| $L \perp\!\!\!\perp T' \| T$ | Yes |
| $L \perp\!\!\!\perp B$ | Yes |
| $L \perp\!\!\!\perp B \| T$ | |
| $L \perp\!\!\!\perp B \| T'$ | |
| $L \perp\!\!\!\perp B \| T, R$ | Yes |