Local Probabilistic Models: Context-Specific CPDs

Sargur Srihari
srihari@cedar.buffalo.edu

Topics

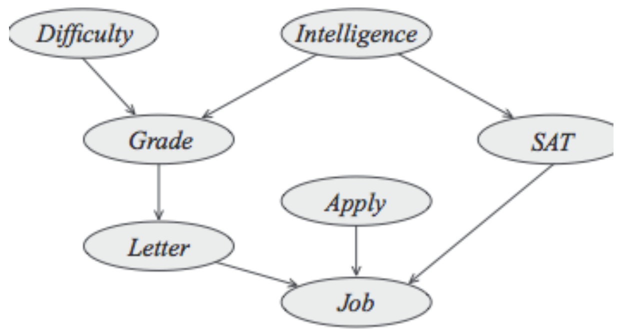
- Context-Specific CPDs
  - Representation
    - Ex: Augmented Student Network
  - Tree CPDs
    - Ex: Printer Diagnosis
  - Rule CPDs
  - Other Representations

Context-Specific CPDs

- Deterministic dependency is one example of structure in CPDs
- A very common type of regularity arises when we have the same effect in several contexts
  - Several different distributions are the same
- Example is given next

Augmented Student BN: Regularity

- Model event: offered a job at Acme Consulting
  - We have a binary variable J
    - j^1: offered job, j^0: otherwise
    - Job depends on SAT & Reco Ltr
    - Student may apply a^1, or not a^0



  - We need to describe CPD: $P(J|A,S,L)$
    - Recruiter is desperate to offer job even without applying

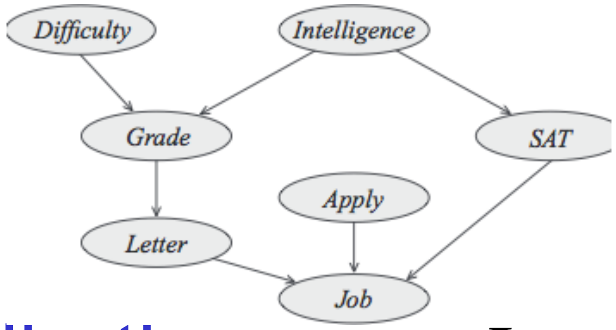
      If $A=a^0$, no access to L and S. So, among 8 values of parents A,S,L, four with $A=a^0$ induces identical distributions over variable J
    - Recruiter feels SAT more important than letter
      - High SAT generates offer without letter: $P(J|a^1,s^1,l^1)=P(J|a^1,s^1,l^0)$
      - Low SAT requires letter
  - Several values of Pa_J specify same conditional probability over J. We need 8 parameters here.

Representing regularity in CPDs

- We have seen several values of Pa_J specify the same conditional probability over J
- How to capture this regularity in our CPD representation
- Many approaches for capturing functions over a scope X that are constant over subsets of instantiations to X
  - Trees
  - Rules

Tree-CPD

- Naturally captures common elements in a CPD
  - Tree for $P(J|A,S,L)$
  - Internal nodes represent tests
    - on parent variables



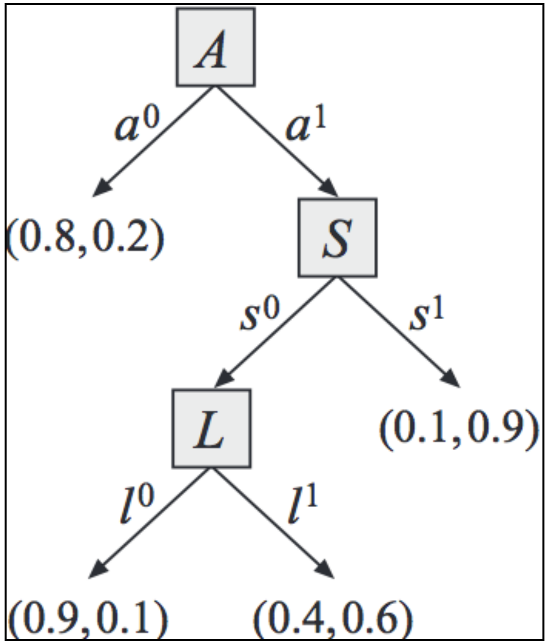
- Leaves are annotated with distribution over J
  - $P(a^0)=0.2$, i.e., probability of offer without applying
  - To determine $P(J|a^1,s^1,l^0)$:

    i.e., student applies, has good SAT

    letter immaterial

    - choose path $A=a^1$ and $S=s^1$
    - $P(j^0)=0.1$, $P(j^1)=0.9$
  - Which is what we use for $P(J|a^1,s^1,l^0)$

    Need 4 parameters instead of 8 in table

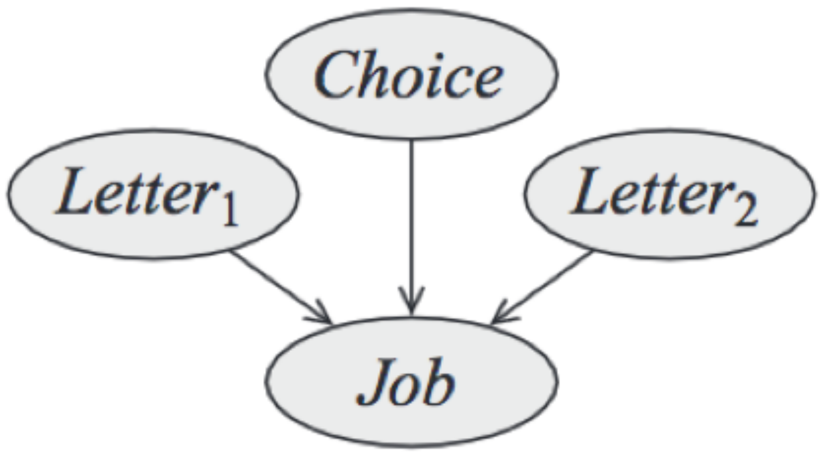


Definition of Tree CPD

- A tree-CPD for a variable X is a rooted tree
- Each t-node in the tree is either a leaf t-node or an interior t-node
- Each leaf is labeled with a distribution $P(X)$
- Each interior node is labeled with some variable $Z \in Pa_X$
- Each interior node has a set of arcs to its children each one associated with a unique assignment $Z = z_i$ for $z_i \in Val(Z)$

Multiplexer CPD

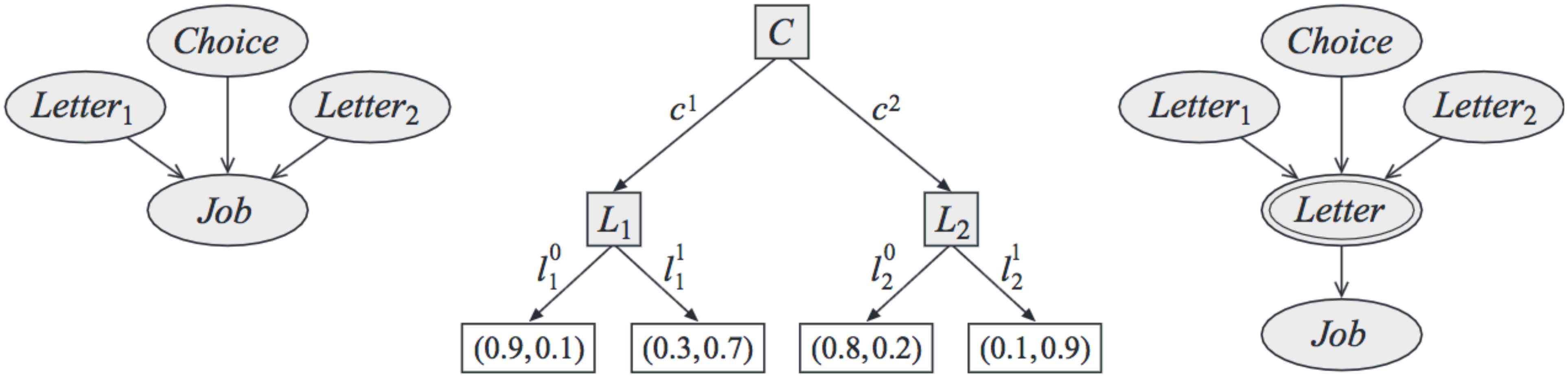
- George has to decide whether to give the recruiter the letter from the Professor of CS 101 or the Professor of CS 102
- Depending on which choice C George makes the dependence will only be on one of the two



Multiplexer CPD

- A CPD $P(Y|A,Z_1,..Z_k)$ is a multiplexer CPD if $Val(A)=\{1,..,k\}$ and $P(Y|a,Z_1,..,Z_k)=\mathbf{1}\{Y=Z_a\}$
  - Where a is the value of A
  - The variable A is called the selector variable of the CPD
- In other words, the value of the selector variable is a copy of the value of one of its parents $Z_1,..,Z_k$
  - The role of A is to select the parent who is being copied

Multiplexer: Tree and BN



(a) network fragment

(b) tree CPD for $P(J|C,L_1,L_2)$

(c) Modified network with new variable L that has a multiplexer CPD

Advantage of Trees

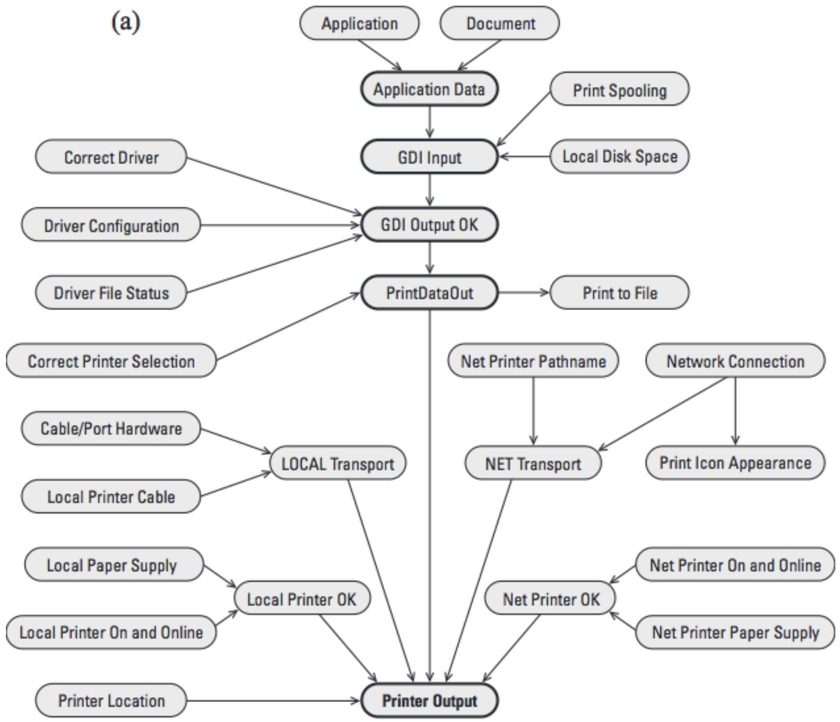
- Provide natural framework for representing context-specificity in a CPD
- People find it convenient
- Lends itself well to automated learning algorithms
  - To construct a tree automatically from a data set

Tree Application: Diagnostic Networks

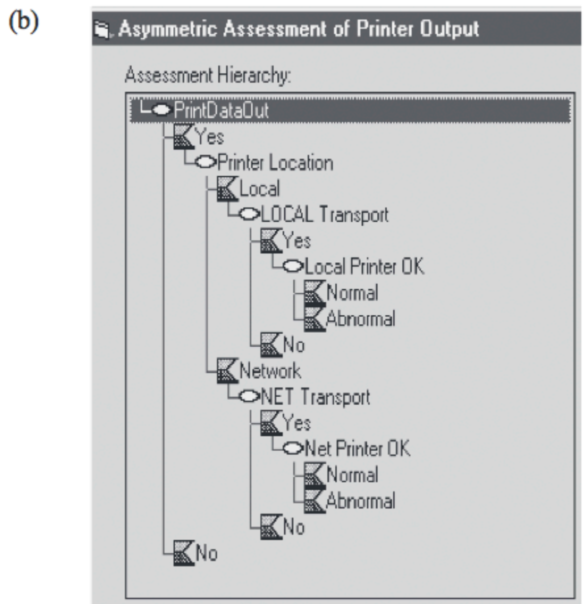
- Trouble-shooting of physical systems
- Context specificity is due to presence of alternative configurations
- Diagnosis of faults in a printer
  - Part of trouble-shooting network for MS Windows 95
  - Printer can be hooked up to either network via
    - Ethernet cable (*Network transport medium*)
      - Affects printer output only if printer is hooked to network
    - Or to local computer via cable (*Local Transport medium)*

Context-Specific Dependencies

(a) Real-world BN for *Microsoft Online Trouble-shooting system*



(b) Structure of Tree-CPD for *Printer Output* variable

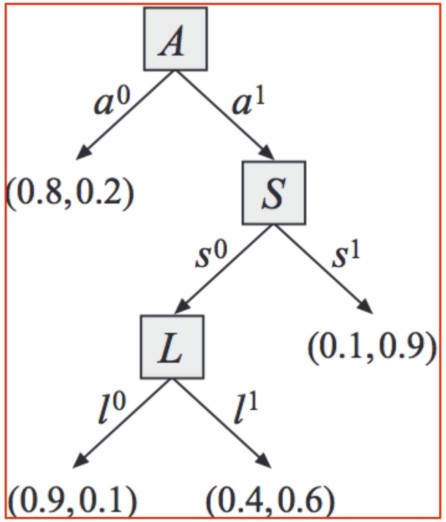


Reduces no. of parameters required from 145 to 55

Rule CPD

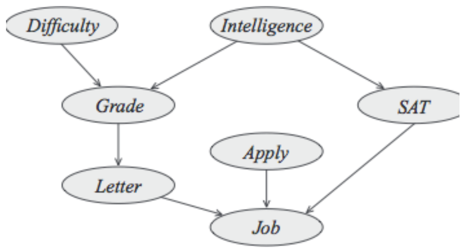
- Trees capture the entire CPD in a single data structure
- A finer-grained specification is via rules
  - Each rule corresponds to a single entry in the CPD of the variable
- A rule ρ is a pair $(c;p)$ where c is an assignment to some subset of variables C and $p \varepsilon [0,1]$.
- C is the scope of ρ denoted $Scope[\rho]$
- This representation decomposes a tree-CPD into its most basic elements

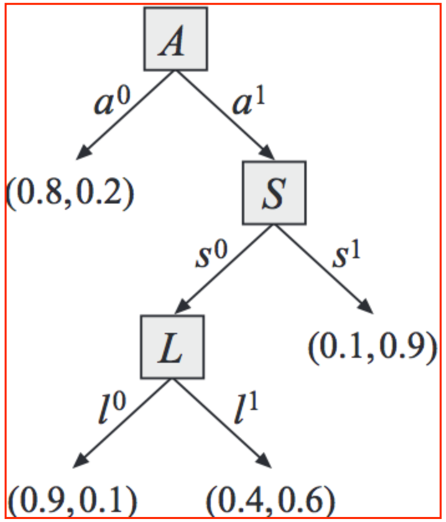
EX: Tree of Augmented Student BN



- There are 8 entries in the CPD tree
- Such that each one corresponds to a branch in the tree and an assignment to the varaible J itself
- Thus the CPD defines eight rules

Ex: CPD of *Job* in Augmented BN





- $\rho_1: <a^0, j^0; 0.8>$
- $\rho_2: <a^0, j^1; 0.2>$
- $\rho_3: <a^1, s^0, l^0, j^0: 0.9>$
- $\rho_3: <a^1, s^0, l^0, j^1: 0.1>$
- $\rho_3: <a^1, s^0, l^0, j^1: 0.4>$
- $\rho_3: <a^1, s^0, l^1, j^1: 0.6>$
- $\rho_3: <a^1, s^1, j^0: 0.1>$
- $\rho_3: <a^1, s^1, j^1: 0.9>$

There are 8 entries in the CPD tree

Such that each one corresponds to a branch in the tree and an assignment to the variable J itself

Thus the CPD defines eight rules

Definition of Rule-based CPD

- A rule-based CPD $p(X|Pa_X)$ is a set of rules R such that
- For each rule ρ ε R we have that $Scope[\rho] \leq \{X\} \vee Pa_X$
- For each assignment $(x,\boldsymbol{u})$ to $\{X\} \vee Pa_X$ we have precisely one rule $(\boldsymbol{c};p)$ ε R such that $\boldsymbol{c}$ is compatible with $(x,\boldsymbol{u})$. In this case we say that $P(X=x|Pa_X=\boldsymbol{u})=p$
- The resulting CPD $P(X|U)$ is a legal CPD in that $\Sigma_x P(x|\boldsymbol{u})=1$

Other Representations

- Tree and rule representations are useful for representation, inference and learning
- However other representations are possible
- They both induce partitions of defined by branches of the tree or rule contexts
- Each partition is associated with a different entry in X's CPD
- Other such methods are decision diagrams, multinets and similarity networks