Valuation-Based Systems

Prakash P. Shenoy

University of Kansas School of Business

Third School on Belief Functions and Their Applications, Stella Plage, France

September 30, 2015

Outline

- Basics of Valuation-Based Systems
- Captain's Problem
- Valuations, Combination, and Marginalization in belief function theory
- Local Computation
- Removal and Conditionals
- Conditional Independence

Outline

- Basics of Valuation-Based Systems
- Captain's Problem
- Valuations, Combination, and Marginalization in belief function theory
- Local Computation
- Removal and Conditionals
- Conditional Independence

Basics of Valuation-based Systems

- A Valuation-based System (VBS) is a formal mathematical system for representation of and reasoning with knowledge.
- It has two parts. A static part that is concerned with representation of knowledge, and a dynamic part that is concerned with reasoning with knowledge
- The static part consists of:
  - Variables: A finite set Φ of variables $f X, Y, Z, \ldots g$. Subsets of Φ will be denoted by $r, s, t, \ldots$.
  - Valuations: A finite set Ψ of valuations $f \rho, \sigma, \tau, \ldots g$. Each valuation encodes knowledge about a subset of variables. Thus, we say, ρ is a valuation for r, where $r \quad \Phi$.
- A graphical representation of a VBS is called a valuation network

Basics of Valuation-based Systems

- The dynamic part consists of several operators:
  - Combination: $: \Psi \quad \Psi$ *!* Ψ that enables us to aggregate knowledge.
  - The combination operator has the following properties:
    - (Domain) If ρ is a valuation for *r*, and σ is a valuation for *s*, then $\rho \quad \sigma$ is a valuation for *r [s*
    - (Commutativity) $\rho \quad \sigma = \sigma \quad \rho$
    - (Associativity) $\rho \quad (\sigma \quad \tau) = (\rho \quad \sigma) \quad \tau$
  - The sequence in which knowledge is aggregated should make no difference.
  - The combination of all valuations, $\quad \Psi$, is called the joint valuation.

Basics of Valuation-based Systems

- Another operator is marginalization
- Marginalization: $-X : \Psi \rightarrow \Psi$ that allows us to coarsen knowledge marginalizing X out of the domain of a valuation.
- Properties of Marginalization
  - (Domain) If ρ is a valuation for r, and $X \in r$, then ρ^{-X} is a valuation for $r \setminus \{X\}$.
  - (Order does not matter) If ρ is a valuation for r, $X \in r$, and $Y \in r$, then $(\rho^{-X})^{-Y} = (\rho^{-Y})^{-X} = \rho^{-\{X,Y\}}$
  - (Local computation) If ρ and σ are valuations for r and s, respectively, $X \in r$, and $X \notin s$, then $(\rho \otimes \sigma)^{-X} = (\rho^{-X}) \otimes \sigma$
- We will sometimes denote $\rho^{-\{X,Y\}}$ by $\rho^{\downarrow r \setminus \{X,Y\}}$

Basics of Valuation-based Systems

- Making inference means finding marginals of the joint valuation for the variables of interest
- Thus, if X is a variable of interest, we compute $(\;\Psi)^{\downarrow X}$ by marginalizing all the other variables in $\Phi \setminus \{X\}$ out of the joint valuation $\;\Psi$.

Basics of Valuation-based Systems

- VBS is an abstraction of several uncertainty calculi
  - propositional calculus
  - probability theory
  - belief function theory
  - Spohn's epistemic belief calculus
  - possibility theory
- VBS can also be considered as an abstraction of
  - optimization using dynamic programming
  - Bayesian decision theory
  - Solving systems of equations
  - relational database theory

Outline

- Basics of Valuation-Based Systems
- Captain's Problem
- Valuations, Combination, and Marginalization in belief function theory
- Local Computation
- Removal and Conditionals
- Conditional Independence

Captain's Problem

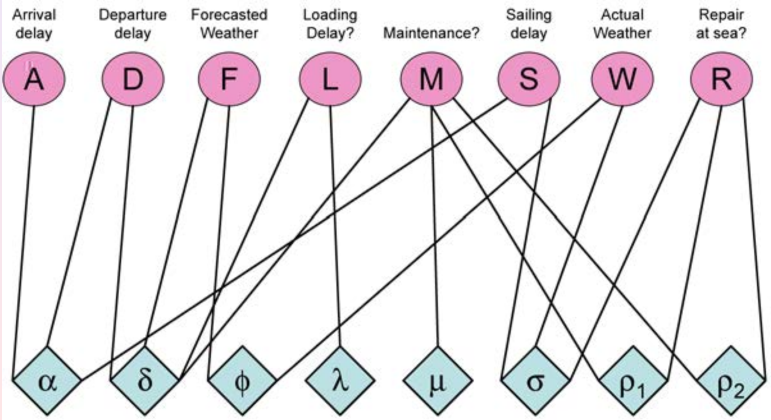
- Captain's Problem (R. Almond, Graphical Belief Modeling, Chapman and Hall, 1995)
  - A ship's captain is concerned about how many days his ship may be delayed before arrival at a destination.
  - The delay in arrival may be a result of delay in departure and/or delay in sailing.
  - Delay in departure may be a result of maintenance (at most 1 day), delay in loading (at most 1 day) or due to forecast of bad weather (at most I day).
  - Delay in sailing may be a result of bad weather (at most 1 day) and/or whether repairs may be needed at sea (at most 1 day).
  - If maintenance is done before sailing, chances of repairs at sea is less likely.
  - Weather forecast says small chance of bad weather (.2), good chance of good weather (0.6). Forecast is 80% reliable.
  - Captain has some knowledge of loading delay, and whether maintenance is done before departure.

Captain's Problem

- Variables
  - A (arrival delay), $\Omega_A = \{0, 1, 2, 3, 4, 5, 6\}$.
  - D (departure delay), $\Omega_D = \{0, 1, 2, 3\}$.
  - S (sailing delay), $\Omega_S = \{0, 1, 2, 3\}$.
  - L (is loading delayed?), $\Omega_L = \{t, f\}$.
  - F (weather forecast), $\Omega_F = \{b, g\}$.
  - W (actual weather), $\Omega_W = \{b, g\}$.
  - M (is maintenance done before sailing?), $\Omega_M = \{t, f\}$.
  - R (is repair at sea needed?), $\Omega_R = \{t, f\}$.

Captain's Problem

- Valuation Network: A bipartite graph with variables and valuations as nodes. Each valuation is linked to the variables in its domain.



Outline

- Basics of Valuation-Based Systems
- Captain's Problem
- Valuations, Combination, and Marginalization in Belief Function Theory
- Local Computation
- Removal and Conditionals
- Conditional Independence

Belief Function Theory

- valuations in belief function theory
  - basic probability assignment (bpa) μ
  - belief function β
  - plausibility function π
  - commonality function χ
- Suppose $s \subseteq \Phi$. The state space of s is $\Omega_s = \times\{\Omega_X | X \in s\}$. Let 2^{Ω_s} denote the set of all subsets of Ω_s. A bpa μ for s is a function $\mu : 2^{\Omega_s} \to [0, 1]$ such that:
  - $\mu(\emptyset) = 0$,
  - $\mu(\mathbf{x}) \geq 0$, and
  - $\Sigma\{\mu(\mathbf{x}) | \mathbf{x} \in 2^{\Omega_s}\} = 1$.
- Subsets $\mathbf{x} \in 2^{\Omega_s}$ such that $\mu(\mathbf{x}) > 0$ are called focal sets of μ.

Belief Function Theory

- A function $\chi : 2^{\Omega_s} \to [0, 1]$ is a commonality function for s if there exists a bpa function μ for s such that

$$\chi(\mathbf{x}) = \Sigma\{\mu(\mathbf{y}) | \mathbf{y} \supseteq \mathbf{x}\}$$

- Suppose μ is a bpa for $\{W\}$ such that $\mu(\{g, b\}) = 1$. Then the corresponding commonality function χ for $\{W\}$ is $\chi(\emptyset) = 1$, $\chi(\{g\}) = 1$, $\chi(\{b\}) = 1$, $\chi(\{g, b\}) = 1$.

Captain's Problem

- Consider the piece of knowledge: Arrival delay is sum of departure delay and sailing delay
- We model this piece of knowledge by a bpa α for fA, D, Sg such that

$$\begin{aligned}\alpha(f(0,0,0),(1,1,0),(2,2,0),(3,3,0),\\ (1,0,1),(2,1,1),(3,2,1),(4,3,1),\\ (2,0,2),(3,1,2),(4,2,2),(5,3,2),\\ (3,0,3),(4,1,3),(5,2,3),(6,3,3)g) = 1.\end{aligned}$$

- α has one focal set. Such bpa are called deterministic.

Captain's Problem

- Loading delay, bad weather forecast, and maintenance each adds one day to departure delay
- We model this piece of knowledge by a bpa δ for $\{D, L, F, M\}$ such that

$$\delta(\{(0, f, g, f), (1, t, g, f), (1, f, b, f), (1, f, g, t), (2, f, b, t), (2, t, g, t), (2, t, g, f), (3, t, b, t)\}) = 1.$$

Captain's Problem

- At least 90% of the time, bad weather and repair at sea each adds 1 day to sailing delay
- We model this by bpa σ for $\{S, A, R\}$ such that

$$\begin{aligned} \sigma(\{(0, g, f), (1, b, f), (1, g, t), (2, b, t)\}) &= 0.9, \\ \sigma(\Omega_{\{S,A,R\}}) &= 0.1 \end{aligned}$$

Captain's Problem

- Forecast is 80% reliable
- This piece of knowledge is represented by bpa ϕ_1 for $\{F, W\}$ such that

$$\begin{aligned}\phi_1(\{(b, b), (g, g)\}) &= 0.8,\\ \phi_1(\Omega_{\{F,W\}}) &= 0.2.\end{aligned}$$

- Forecast predicts bad weather with chance 0.2 and good weather with chance 0.6
- This piece of knowledge is represented by bpa ϕ_2 for $\{F\}$ such that

$$\begin{aligned}\phi_2(\{b\}) &= 0.2,\\ \phi_2(\{g\}) &= 0.6,\\ \phi_2(\Omega_{\{F\}}) &= 0.2.\end{aligned}$$

Captain's Problem

- Loading is delayed with chance 0.3, and on schedule with chance 0.5.
- This piece is model by bpa λ for $\{L\}$ such that

$$\begin{aligned}\lambda(\{t\}) &= 0.3,\\ \lambda(\{f\}) &= 0.5,\\ \lambda(\Omega_{\{L\}}) &= 0.2.\end{aligned}$$

- No maintenance was done on the ship prior to departure
- This piece of knowledge is represented by bpa μ for $\{M\}$ such that

$$\mu(\{f\}) = 1.$$

Captain's Problem

- Consider valuation ρ_1 for fM, Rg as follows (If maintenance was done prior to sailing, then chances of repair at sea is between 10 and 30%)

$$
\begin{aligned}
\rho_1(f(t,t),(f,t),(f,f)g) &= 0.1,\\
\rho_1(f(t,f),(f,t),(f,f)g) &= 0.7,\\
\rho_1(\Omega_{\{M,R\}}) &= 0.2.
\end{aligned}
$$

- Consider valuation ρ_2 for fM, Rg as follows (If maintenance was not done prior to sailing, then chances of repair at sea is between 20 and 80%)

$$
\begin{aligned}
\rho_2(f(f,t),(t,t),(t,f)g) &= 0.2,\\
\rho_2(f(f,f),(t,t),(t,f)g) &= 0.2,\\
\rho_2(\Omega_{\{M,R\}}) &= 0.6.
\end{aligned}
$$

Belief Function Theory

- Combination in belief function theory is Dempster's rule of combination, also called product-intersection rule:
  - The product of the bpa values is assigned to the intersection of the subsets, The probability assigned to the ; is reduced to 0, and the remaining probabilities are normalized so they sum to 1.
- In terms of commonality functions, Dempster's rule of combination is pointwise multiplication followed by normalization.

Captain's Problem

- Suppose μ is a bpa for M that represents that maintenance was not done prior to sailing, i.e.,

$$\mu(\{f\}) = 1 \Rightarrow \mu^{\{M,R\}}(\{(f,t),(f,f)\}) = 1$$

- Now consider $\mu \quad \rho_1$

$\mu \quad \rho_1$	$\{(t,t),(f,t),(f,f)\}$ 0.1	$\{(t,f),(f,t),(f,f)\}$ 0.7	$\Omega_{\{M,R\}}$ 0.2
$\{(f,t),(f,f)\}$ 1	$\{(f,t),(f,f)\}$ 0.1	$\{(f,t),(f,f)\}$ 0.7	$\{(f,t),(f,f)\}$ 0.2

i.e.,$(\mu \quad \rho_1)(\{(f,t),(f,f)\} = 1$

Captain's Problem

- Now consider $\mu \quad \rho_2$

$\mu \quad \rho_2$	$f(f,t),(t,t),(t,f)g$ 0.2	$f(f,f),(t,t),(t,f)g$ 0.2	$\Omega_{\{M,R\}}$ 0.6
$f(f,t),(f,f)g$ 1	$f(f,t)g$ 0.2	$f(f,f)g$ 0.2	$f(f,t),(f,f)g$ 0.6

$$\begin{aligned} \text{i.e.,}(\mu \quad \rho_2)(f(f,t)g) &= 0.2, \\ (\mu \quad \rho_2)(f(f,f)g) &= 0.2, \\ (\mu \quad \rho_2)(f(f,t),(f,f)g) &= 0.6 \end{aligned}$$

Belief Function Theory

- Dempster's rule satisfies all properties of combination
  - (Domain) If μ_1 is a bpa for s_1 and μ_2 is a bpa for s_2, then $\mu_1 \quad \mu_2$ is a bpa for s_1 [s_2
  - (Associativity) $\mu_1 \quad (\mu_2 \quad \mu_3) = (\mu_1 \quad \mu_2) \quad \mu_3$
  - (Commutativity) $\mu_1 \quad \mu_2 = \mu_2 \quad \mu_1$

Belief Function Theory

- Marginalization in belief function theory is addition.
- Projection of States: If **a** $2\ \Omega_s$, and $X\ 2\ s$, then $\mathbf{a}^{\downarrow s \setminus \{X\}} = \mathbf{a}^{-X}$ is the state of $s\ n\ f X g$ obtained from **a** by dropping the state of X.
- If μ is a bpa for s, and $X\ 2\ s$, then μ^{-X} is defined as follows:

$$\mu^{-X}(\mathbf{a}) = \Sigma f \mu(\mathbf{y}) j \mathbf{y}^{-X} = \mathbf{a} g$$

Captain's Problem

- Consider ρ_1 for $\{M, R\}$

$$\begin{aligned}\rho_1(\{(t,t),(f,t),(f,f)\}) &= 0.1,\\ \rho_1(\{(t,f),(f,t),(f,f)\}) &= 0.7,\\ \rho_1(\Omega_{\{M,R\}}) &= 0.2.\end{aligned}$$

- Then ρ_1^{-M} for $\{R\}$ is as follows:

$$\rho^{-M}(\{t, f\}) = 1.$$

- And ρ_1^{-R} for $\{M\}$:

$$\rho^{-R}(\{t, f\}) = 1.$$

- Thus, ρ_1 by itself tells us nothing about M or R.

Captain's Problem

- Now consider $\mu \quad \rho_1$ for $f M, Rg$:

$$(\mu \quad \rho_1)(f(f, t), (f, f)g) = 1.$$

- Then, $(\mu \quad \rho_1)^{-M}$ for $f Rg$ is given by:

$$(\mu \quad \rho_1)(f t, f g) = 1.$$

i.e., $(\mu \quad \rho_1)^{-M}$ tells us nothing about R.

Captain's Problem

- Now consider $\mu \quad \rho_2$ for fM, Rg:

$$
\begin{aligned}
(\mu \quad \rho_2)(f(f, t)g) &= 0.2, \\
(\mu \quad \rho_2)(f(f, f)g) &= 0.2, \\
(\mu \quad \rho_2)(f(f, t), (f, f)g) &= 0.6.
\end{aligned}
$$

- Then, $(\mu \quad \rho_2)^{-M}$ for fRg is given by:

$$
\begin{aligned}
(\mu \quad \rho_2)^{-M}(ftg) &= 0.2, \\
(\mu \quad \rho_2)^{-M}(ffg) &= 0.2, \\
(\mu \quad \rho_2)^{-M}(ft, fg) &= 0.6.
\end{aligned}
$$

Belief Function Theory

- The definition of marginalization of bpa function satisfies the properties of marginalization:
  - (Domain) If ρ is a bpa for r, and $X \in r$, then ρ^{-X} is a bpa for $r \setminus \{X\}$.
  - (Order does not matter) If ρ is a bpa for r, $X \in r$, and $Y \in r$, then $(\rho^{-X})^{-Y} = (\rho^{-Y})^{-X} = \rho^{-\{X,Y\}} = \rho^{\downarrow r \setminus \{X,Y\}}$.
  - (Local computation) If ρ and σ are bpa's for r and s, respectively, $X \in r$, and $X \notin s$, then $(\rho \quad \sigma)^{-X} = (\rho^{-X}) \quad \sigma$.

Outline

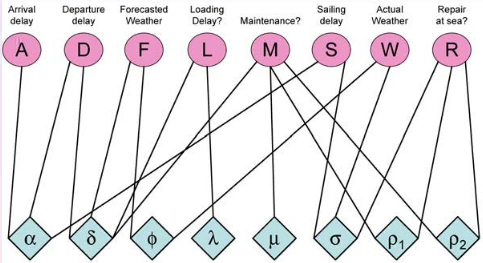
- Basics of Valuation-Based Systems
- Captain's Problem
- Valuations, Combination, and Marginalization in Belief Function Theory
- Local Computation
- Removal and Conditionals
- Conditional Independence

Local Computation

- Making inference means finding marginals of the joint valuation $\otimes\Psi$ for the variables of interest.
- If there are many variables in Φ, computing the joint valuation $\otimes\Psi$ for Φ is intractable.
- However, one can compute the marginal of the joint for X, $(\otimes\Psi)^{\downarrow X}$, without computing the joint explicitly, using so-called local computation.
- The axiom that allows local computation is the local computation axiom:
  If ρ and σ are bpa's for r and s, respectively, $X \in r$, and $X \notin s$, then $(\rho \otimes \sigma)^{-X} = (\rho^{-X}) \otimes \sigma$.

Local Computation

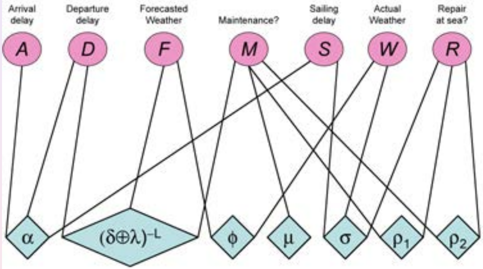
- Consider the Captain's problem. We would like to compute the marginal of the joint for *A*. So we have to marginalize all other variables from the joint.



- Consider *L*. It is in the domain of δ and λ only. The local computation axiom guarantees that if we replace δ and λ by $(\delta \quad \lambda)^{-L}$, then the product of all valuations will give us $(\quad \Psi)^{-L}$.

Local Computation

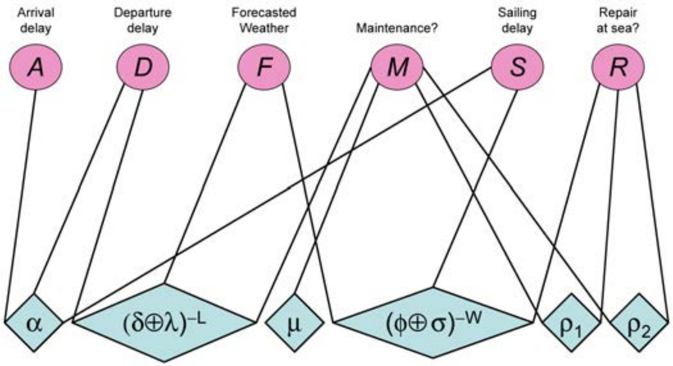
- The reduced VN is as follows:



- We can recursively remove all but *A* from the VN.
- Consider *W*. It is in the domain of ϕ for $\{F, W\}$ and σ for $\{S, W, R\}$. Thus, $(\phi \quad \sigma)^{-W}$ will be a bpa for $\{F, S, R\}$.

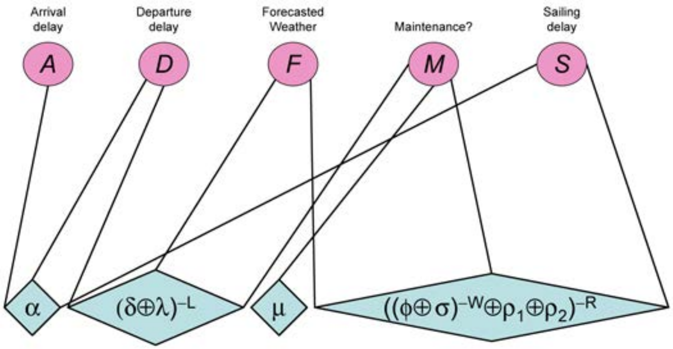
Local Computation

- After deletion of $\{L, W\}$:



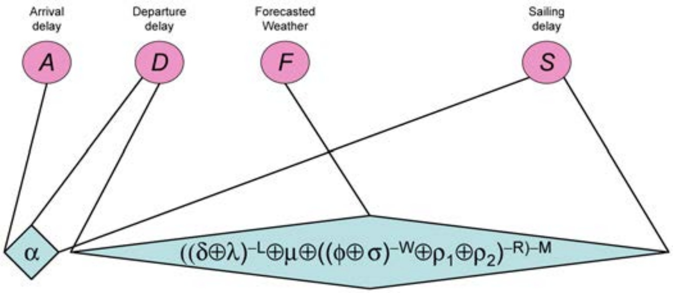
Local Computation

- After deletion of fL, W, Rg:



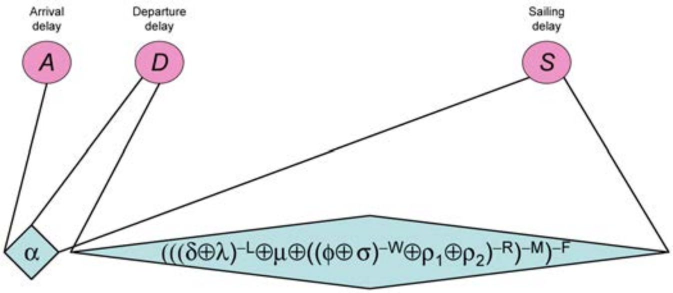
Local Computation

- After deletion of $f L, W, R, M g$:



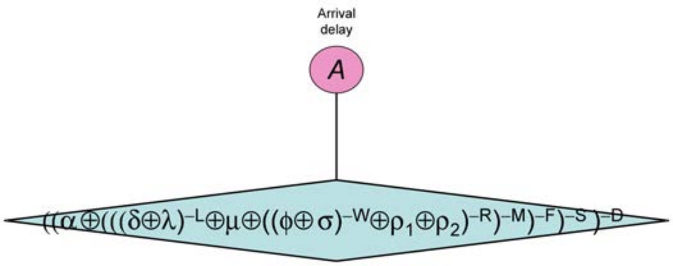
Local Computation

- After deletion of $f L$, W, R, M, Fg:



Local Computation

- After deletion of $f L, W, R, M, F, S, Dg$:



- If we combine all valuations, we have the marginal of the joint for A.

Local Computation

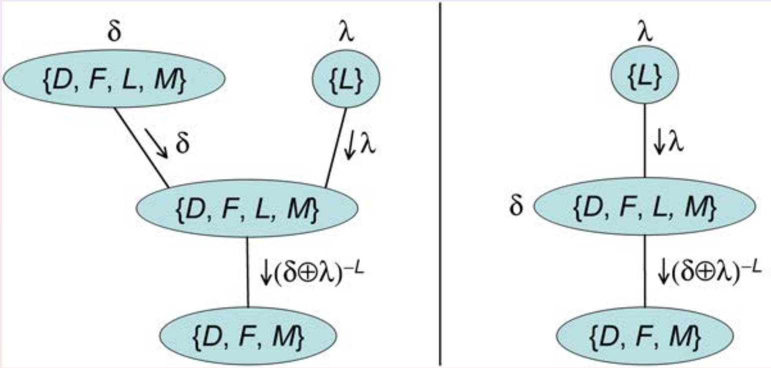
- In finding the marginal for A, we used deletion sequence $L\ W\ R\ M\ F\ S\ D$.
- The order does not matter axiom allows us to use any deletion sequence (and obtain the same marginal).
- Some deletion sequences involve less computation than others.
- Finding an optimal deletion sequence is a hard problem.
- So we use heuristics to select a sequence such as one-step-look-ahead: The variable to be marginalized next is the one that leads to combination on the smallest domain.

Local Computation

- If we can find the marginal for A, we can find the marginal for any other variable in a similar manner.
- However, there may be duplication of computations.
- We can avoid duplication by saving intermediate computations in a data structure called a “join tree.”
- A join tree is a tree with subsets of variables as nodes having the property that if a variable appears in two different nodes, then it appears in all nodes in the path between the two nodes.

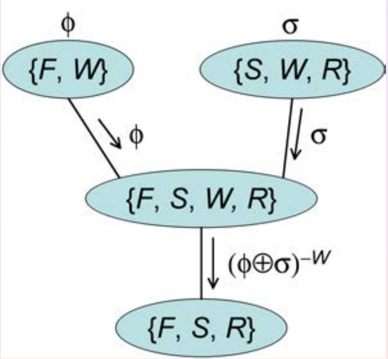
Local Computation

- Consider deletion of L. We can describe the computation as messages between nodes as follows:



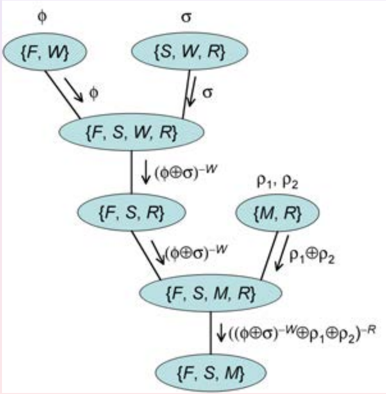
Local Computation

- Similarly, deletion of W can be described as follows:



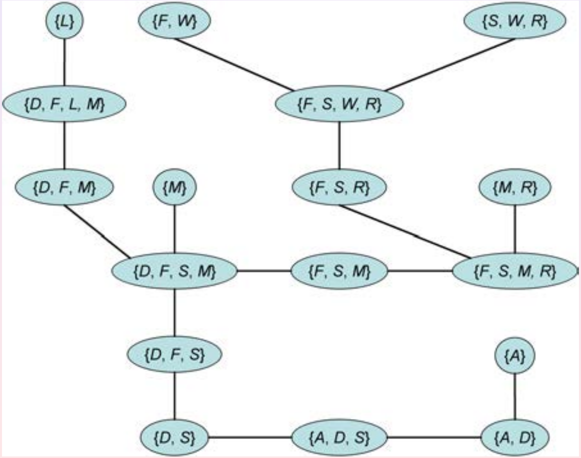
Local Computation

- Similarly, deletion of *R* can be described as follows:



Local Computation

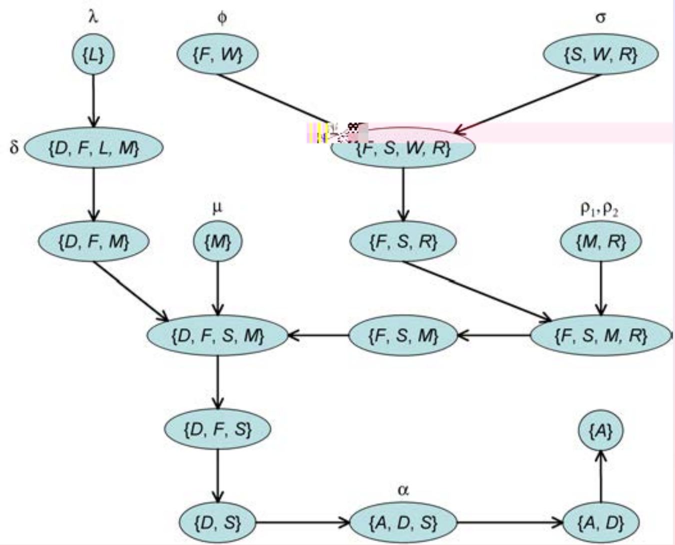
- The tree thus constructed is a join tree.



Local Computation

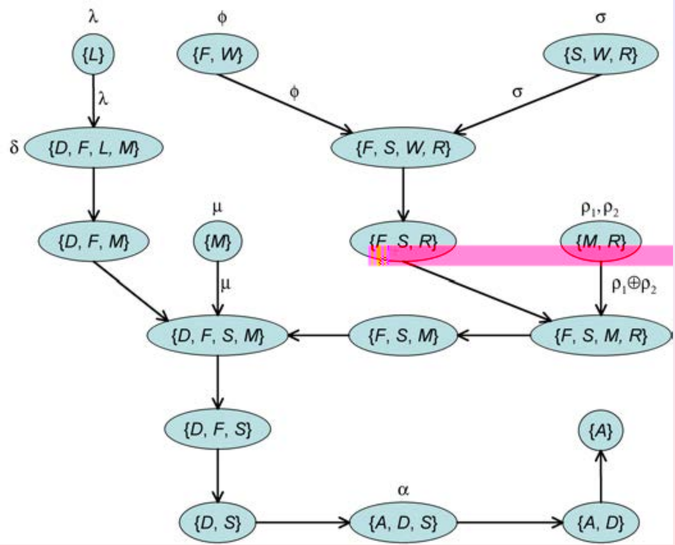
- To find the marginal for a variable, say A, orient all edges toward A, which is now the root of the tree.
- Each node sends a message to its inward neighbor, which is the combination of all messages it receives from its outward neighbors plus what it has suitably marginalized.
- Timing: Leaves (nodes with no outward neighbors) can send messages right away. Non-leaves have to wait till they receive a message from all its outward neighbors.
- Process is finished when the root has received a message from its outward neighbors. The root then combines all the messages it receives plus what it has.

- At the beginning:

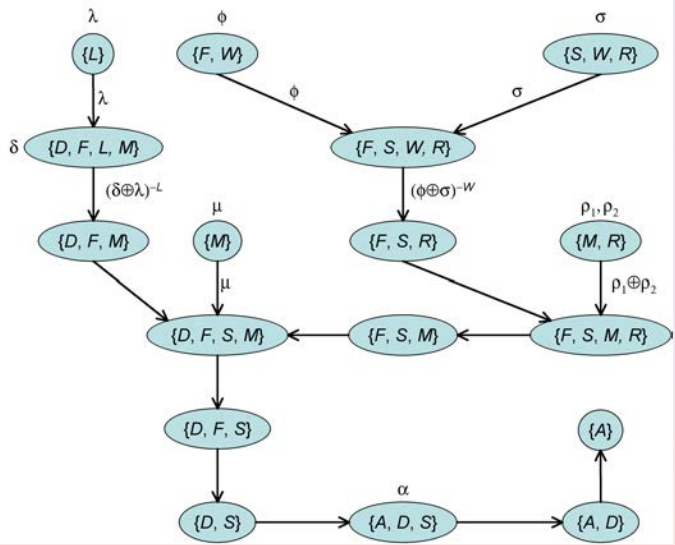


Local Computation

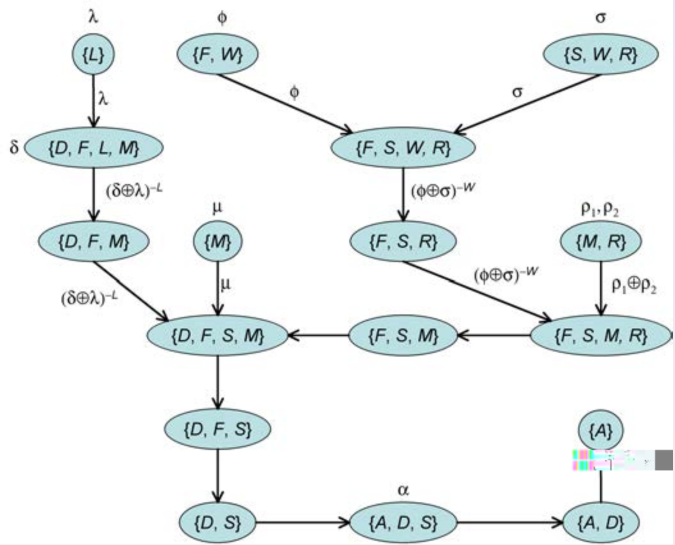
- Time step 1:



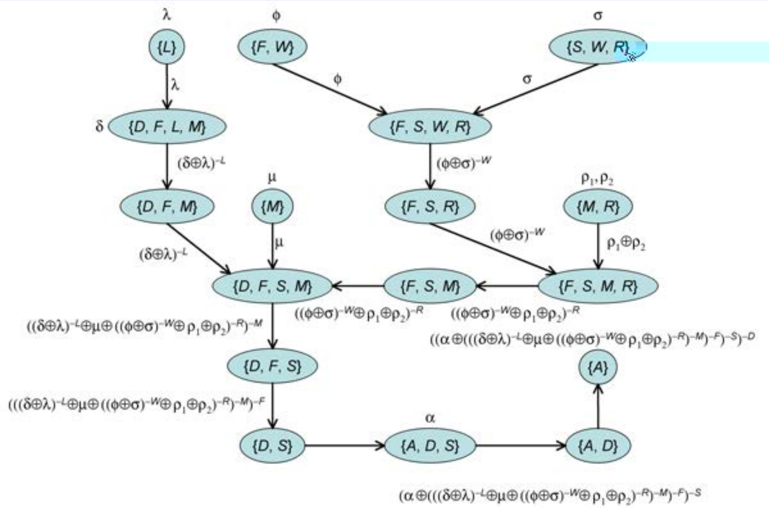
- Time step 2:


λ
ϕ
σ
{L}
{F, W}
{S, W, R}
λ
ϕ
σ
δ
{D, F, L, M}
{F, S, W, R}
(δ⊕λ)^{-L}
(ϕ⊕σ)^{-W}
μ
ρ1, ρ2
{D, F, M}
{M}
{F, S, R}
{M, R}
μ
ρ1⊕ρ2
{D, F, S, M}
{F, S, M}
{F, S, M, R}
{D, F, S}
{A}
α
{D, S}
{A, D, S}
{A, D}

- Time step 3:


λ
{L}
λ
ϕ
{F, W}
ϕ
σ
{S, W, R}
σ
δ
{D, F, L, M}
{F, S, W, R}
(δ⊕λ)^{-L}
(ϕ⊕σ)^{-W}
μ
{D, F, M}
{M}
{F, S, R}
ρ1, ρ2
{M, R}
μ
(ϕ⊕σ)^{-W}
ρ1⊕ρ2
(δ⊕λ)^{-L}
{D, F, S, M}
{F, S, M}
{F, S, M, R}
{D, F, S}
{A}
α
{D, S}
{A, D, S}
{A, D}

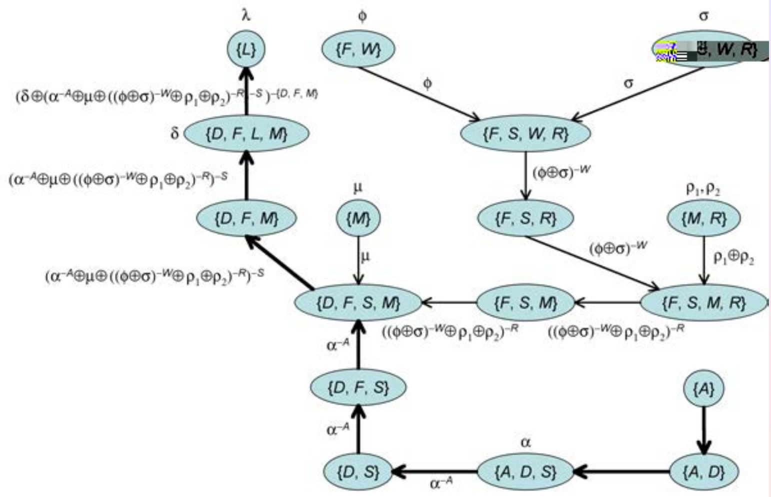
- End of propagation for *A*:



Local Computation

- Suppose we now wish to find the marginal for L.
- We now re-orient some of the edges so L is the root, and compute new messages for the re-oriented edges.

- End of propagation for *L*:



Outline

- Basics of Valuation-Based Systems
- Captain's Problem
- Valuations, Combination, and Marginalization in Belief Function Theory
- Local Computation
- Removal and Conditionals
- Conditional Independence

Removal and Conditionals

- We now define a removal operator, which we will use to define conditionals.
- Removal $: \Psi \quad \Psi \; ! \; \Psi$ that allows us to remove one valuation from another
- The removal operator satisfies the following axioms:
  - (Domain): Suppose σ is a valuation for *s* and ρ is a valuation for *r*. Then $\sigma \quad \rho$ is a valuation for *r* [*s*.
  - (Identity): For each valuation ρ for *r*, there exists an identity valuation ι_ρ for *r* such that $\rho \quad \rho = \iota_\rho$
  - (Combination and Removal): $(\pi \quad \theta) \quad \rho = \pi \quad (\theta \quad \rho)$.
- We call $\sigma \quad \rho$, read as σ minus ρ, the valuation resulting after removing ρ from σ

Removal and Conditionals

- In belief function theory, removal is pointwise division of commonality functions followed by normalization.
- Suppose σ is a commonality function for s, and suppose ρ is a commonality function for r. Then, $\sigma \ominus \rho$ is a valuation for $r \cup s$ given by:

$$(\sigma \ominus \rho)(\mathbf{x}) = \begin{cases} K^{-1}\sigma(\mathbf{x}^{\downarrow s})/\rho(\mathbf{x}^{\downarrow r}) & \text{if } K > 0 \text{ and } \rho(\mathbf{x}^{\downarrow r}) > 0 \\ 0 & \text{if } K = 0 \text{ or } \rho(\mathbf{x}^{\downarrow r}) = 0 \end{cases}$$

where

$$K = \Sigma\{(-1)^{|\mathbf{x}|+1}\sigma(\mathbf{x}^{\downarrow s})/\rho(\mathbf{x}^{\downarrow r}) \mid \mathbf{x} \in 2^{\Omega_{r\cup s}} \text{ such that } \rho(\mathbf{x}^{\downarrow r}) > 0\}$$

Removal and Conditionals

- Suppose τ is a joint valuation for Φ, and suppose r and s are disjoint subsets of Φ. We will call $\tau^{\downarrow r \cup s} \quad \tau^{\downarrow r}$ the conditional for s given r with respect to τ, and denote it by $\tau(s\ j\ r)$
- If $r = \ ;$, let $\tau(s)$ denote $\tau(s\ j\ ;)$
- Properties of Conditionals
  - $\tau(s) = \tau^{\downarrow s}$
  - $\tau(r) \quad \tau(s\ j\ r) = \tau(r\ [\ s)$
  - $\tau(r\ j\ t) \quad \tau(s\ j\ r\ [\ t) = \tau(r\ [\ s\ j\ t)$
  - $\tau(s\ j\ r)^{\downarrow r} = \iota_{\tau(r)}$, i.e., $\tau(r) \quad (\tau(s\ j\ r)^{\downarrow r}) = \tau(r)$.
  - etc.

Outline

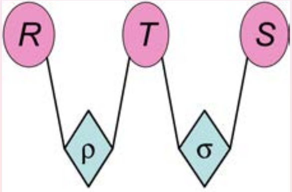
- Basics of Valuation-Based Systems
- Captain's Problem
- Valuations, Combination, and Marginalization in Belief Function Theory
- Local Computation
- Removal and Conditionals
- Conditional Independence

Conditional Independence

- DEFINITION: Suppose τ denotes the joint valuation, and suppose *r*, *s*, and *t* are disjoint subsets of variables. We say *r* is conditionally independent of *s* given *t* with respect to τ, written as $r\,?_{\tau} s \; j \; t$, if there exists ρ for $r \; [\; t$ and σ for $s \; [\; t$ such that

$$\tau(r \; [\; s \; [\; t) = \rho \quad \sigma$$

- For example, in the model below, $f R g \, ?_{\tau} \, f S g \; j \; f T g$, where $\tau = \rho \quad \sigma$

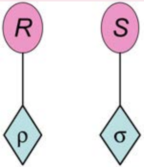


Conditional Independence

- A special case of conditional independence is independence when $t = ;$
- DEFINITON: Suppose τ denotes the joint valuation, and suppose r, and s are disjoint subsets of variables. We say r is independent of s with respect to τ, written as $r ?_\tau s$, if $r ?_\tau s \; j \; ;$, i.e., there exists ρ for r and σ for s such that

$$\tau(r \;[\; s) = \rho \quad \sigma$$

- For example, in the model below, $f R g ?_\tau f S g$, where $\tau = \rho \quad \sigma$

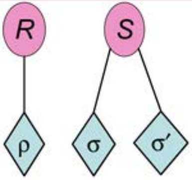


Conditional Independence

- What does it mean to say $fRg\,?_{T}\,fSg$?
- There are several semantics of independence
  - Irrelevance
  - Factorization
  - No double counting of evidence (when double counting matters)

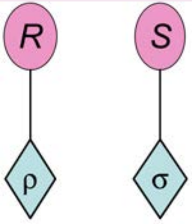
Conditional Independence

- Independence in terms of irrelevance
- In probability theory, $f R g ?_P f S g$ if $P(R \; j \; s_1) = P(R)$, for all $s_1 \; 2 \; \Omega_S$.
- Suppose $f R g ?_\tau f S g$. If we get some evidence about S, denoted say by σ' for $f S g$, then the marginal for R remains unchanged, i.e., if $\tau' = \rho \quad \sigma \quad \sigma'$, then $(\tau')(f R g) = (\rho \quad \sigma \quad \sigma')^{-S} = \rho \quad (\sigma \quad \sigma')^{-S} = \rho = \tau(f R g)$



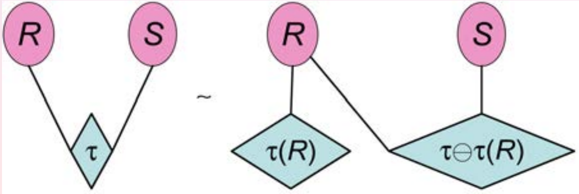
Conditional Independence

- Independence in terms of factorization
- In probability theory, $fRg ?_f fSg$ if $f(r, s) = g(r)h(s)$, where f is the joint PDF of R and S, r 2 Ω_R, and s 2 Ω_S.
- Suppose $fRg ?_\tau fSg$. Then $\tau = \rho \quad \sigma$ factors into a valuation ρ for fRg and a valuation σ for fSg.



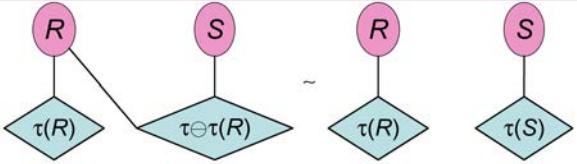
Conditional Independence

- Independence in terms of No double counting (when double counting matters).
- In general, $\sigma \oplus \sigma \neq \sigma$. So, we should be careful to not count a piece of evidence more than once.
- Suppose τ is a valuation for $\{R, S\}$. We can always factor τ as follows: $\tau(R)$ and $\tau(S \mid R) = (\tau \ominus \tau(R))$ and there is no double-counting.



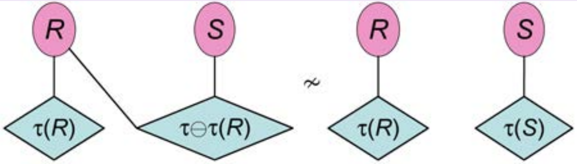
Conditional Independence

- If $f R g ?_{\tau} f S g$, then $\tau \quad \tau(R) = (\tau(R) \quad \tau(S)) \quad \tau(R) = \tau(S) \quad \iota_{\tau(R)}$. Thus, if we combine $\tau(R)$ and $\tau(S)$, we get τ, and there is no double counting.



Conditional Independence

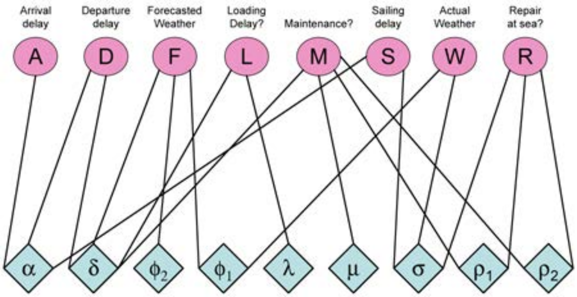
- If $f R g$ and $f S g$ are not independent, then $\tau(R) \quad \tau(S) \neq \tau$, since there is double-counting of evidence.



- $\tau = \tau(R) \quad (\tau \quad \tau(R))$. So $\tau(S) = (\tau(R) \quad (\tau \quad \tau(R)))^{\downarrow\{S\}}$. Thus, in combining $\tau(R)$ and $\tau(S)$, we are counting the information in $\tau(R)$ twice.
- Dempster's rule of combination should only be used to combine belief functions that are "distinct" or "independent", i.e., there is no double counting.

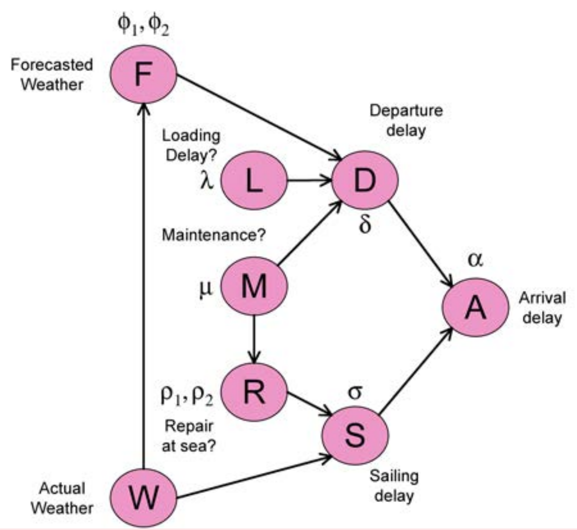
Conditional Independence

- One set of assumptions in the Captain's problem is that there is no double counting in the valuations in the model.



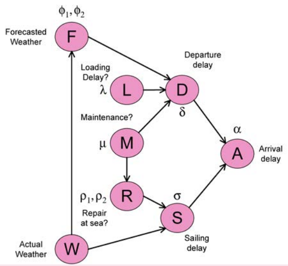
Conditional Independence

- The Captain problem can also be described by a causal DAG as follows:



Conditional Independence

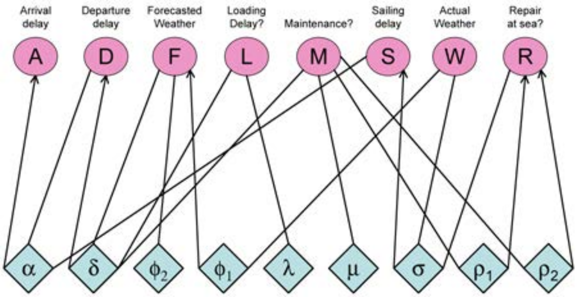
- We can deduce the following set of conditional independence assumptions from the causal model:



$$\{L\} \perp \{M\}, \quad \{W\} \perp \{L, M\}, \quad \{F\} \perp \{L, M\} \mid \{W\},$$
$$\{R\} \perp \{L, W, F\} \mid \{M\}, \quad \{D\} \perp \{W, S\} \mid \{F, L, M\},$$
$$\{S\} \perp \{L, M, F, D\} \mid \{R, W\}, \quad \{A\} \perp \{L, M, W, F, R\} \mid \{D, S\}$$

Conditional Independence

- We can derive the same set of conditional independencies from the VN if we include the fact that some of the valuations are conditionals (denoted by directed edges) as shown below.



References

- VBS: Shenoy, PP, “A Valuation-Based Language for Expert Systems,” *Int. J. of Approx. Reas.*, 3(2), 1989, 383–411
- Captain’s Problem: Almond, R, *Graphical Belief Modeling*, 1995, Chapman & Hall
- Local Computation: Shenoy, PP & G Shafer, “Axioms for Probability and Belief-Function Propagation,” in Shachter, RD, T Levitt, JF Lemmer and LN Kanal (eds.), *Uncert. in Art. Int.*, 4, 1990, 169–198
- Conditional Independence: Shenoy PP, “Conditional Independence in Valuation-Based Systems,” *Int. J. of Approx. Reas.*, 10(3), 1994, 203–234
- Valuation Networks: Shenoy, PP, “Representing Conditional Independence Relations by Valuation Networks,” *Int. J. of Uncert., Fuzziness & Knowledge-Based Sys.*, 2(2), 1994, 143–165

Questions

Any Questions?

Exercises

1. Explain why it is important to propagate in a join tree (see slide # 41) to get the correct marginals? Provide a simple example to demonstrate than propagation in a tree of subsets that is not a join tree will result in incorrect marginals
2. For the Captain's Problem, what is the marginal of the joint for *M*? Provide an answer symbolically similar to the marginal for *A* shown in slide #39.
3. Consider the case where we have valuations α for fAg, β for B, and a conditional valuation χ for C given A and B. Let τ denote the joint valuation for fA, B, Cg. Show that A and B are independent with respect to $\tau(fA, Bg) = \tau^{\downarrow\{A,B\}}$, but that this independence may be lost on observing $C = c$.