Probabilistic Aggregate Models

Motivation

- Models for scene interpretation can be conveniently strctured in compositional and taxonomical hierarchies.
- Properties of aggregate models are often probabilistic in nature
- Probabilities can provide guidance for interpretation steps

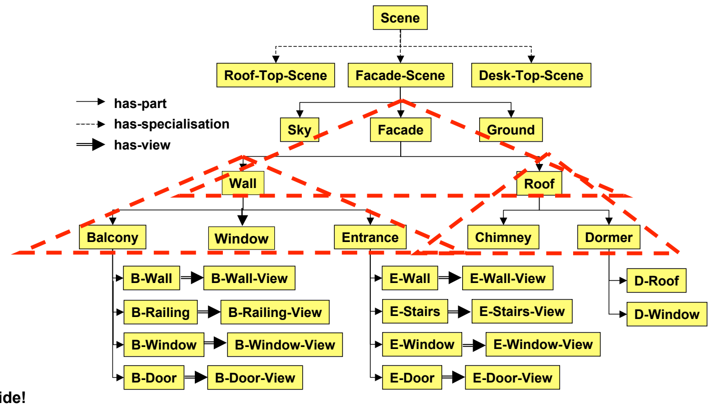
=> How can one realize probabilistic inferences based on aggregate descriptions in a compositional hierarchy?

Animated Slide!

High-level Knowledge Structure

To interface with human concepts and common knowledge, a generic approach requires:

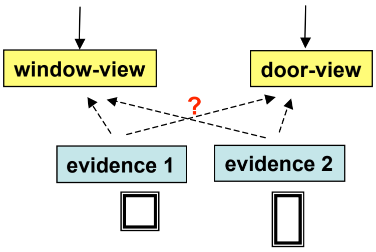
- Object-centered representations
- Compositional hierarchies with abstraction => aggregates



Animated Slide!

Evidence Assignment Problem in Facade Domain

To which part of an aggregate should a given evidence be assigned?

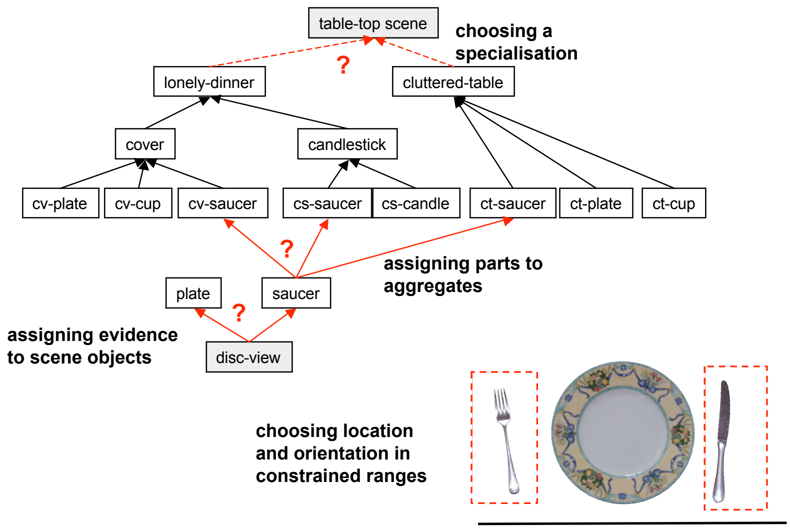


Optimal decision would require

- postponing classification until all evidence is available
- maximization over all reasonable evidence permutations

Assignment problem not encountered in Bayesian decisions or belief system reasoning!

Uncertain Decisions in Table-Setting Domain

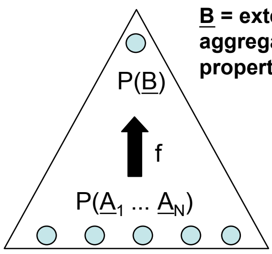


Frequentist Probabilistic Model

<u>Basic view</u>:

An aggregate

- **is a set of parts which tend to co-occur probabilistically and together constitute a meaningful entity**
- **specifies an abstraction from the descriptions of its parts**



B = external aggregate properties

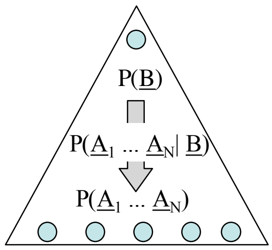
Example: Bounding-box abstraction

$\underline{A}_1$ ... $\underline{A}_N$ = internal parts properties

There exists a functional mapping f : $\underline{A}_1$ ... $\underline{A}_N$ => $\underline{B}$

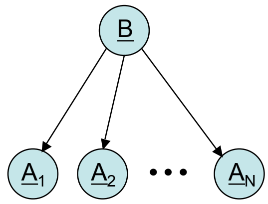
Probabilistic Aggregate Structure

external representation in terms of aggregate properties



internal representation in terms of component properties

Rimey 93:
Tree-shaped part-of nets, is-a trees, expected-area nets, and task nets



⇒ **unrealistic conditional independence:**

$P(\underline{A}_1 \ldots \underline{A}_N | \underline{B}) = P(\underline{A}_1|\underline{B})\, P(\underline{A}_2|\underline{B}) \ldots P(\underline{A}_N|\underline{B})$

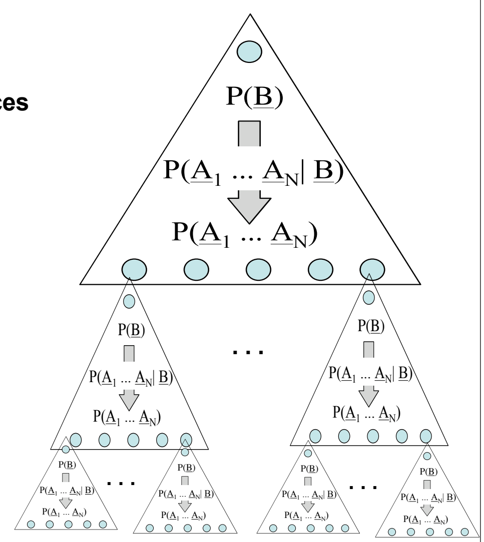
Probabilistic Aggregate Hierarchy

What are useful (and plausible) independence assumptions

- **for efficient probabilistic inferences**
- **for intuitive aggregate models?**

Simplifying assumptions (initially):

- **Distinct names for multiple parts of the same kind**
- **Fixed set of parts per aggregate**
- **No specialization branchings**



Bayesian Compositional Hierarchy (1)

Conditional-independence requirements for a compositional hierarchy to be an "Bayesian compositional hierarchy":

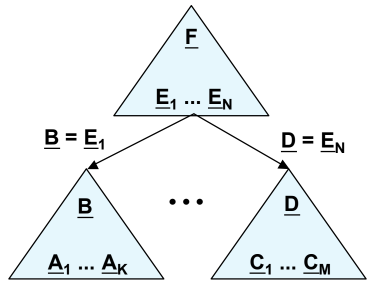
$\underline{X}$	an aggregate node
$\underline{Y}_1 .. \underline{Y}_N$	the parts of $\underline{X}$
succ($\underline{X}$)	all successors of $\underline{X}$

Req 1: $P(\underline{X} \mid succ(\underline{X})) = P(\underline{X} \mid \underline{Y}_1 .. \underline{Y}_N)$ **(1)**

Aggregate properties do not depend on details below the part properties.

Example:

Given $\underline{E}_1 ... \underline{E}_N$, F is independent of all successors below $\underline{E}_1 ... \underline{E}_N$



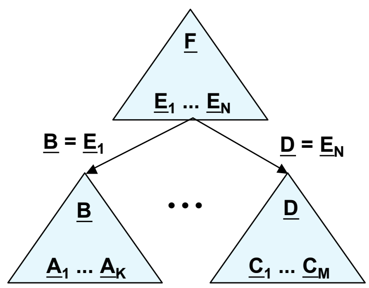
Bayesian Compositional Hierarchy (2)

Req 2: $P(succ(\underline{Y}_i) \mid \underline{Y}_1 .. \underline{Y}_N) = P(succ(\underline{Y}_i) \mid \underline{Y}_i)$ **(2)**

Part properties depend only on the properties of the corresponding mother aggregate.

Example:

Given $\underline{B} = \underline{E}_1$, $\underline{A}_1 ... \underline{A}_K$ and their successors are independent of $\underline{E}_2 ... \underline{E}_N$



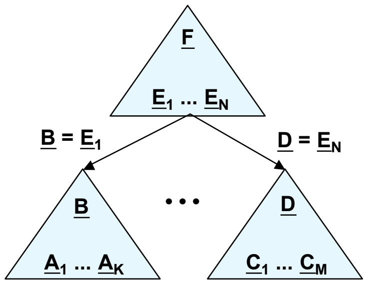
Bayesian Compositional Hierarchy (3)

Req 3: $P(succ(\underline{Y}_1 .. \underline{Y}_N) \mid \underline{Y}_1 .. \underline{Y}_N) = \Pi\ P(succ(\underline{Y}_i) \mid \underline{Y}_1 .. \underline{Y}_N)$ (3)

Parts of different aggregates are statistically independent given their mother aggregates.

Example:

Given $\underline{E}_1 ... \underline{E}_N$, $\underline{A}_1 ... \underline{A}_K$ and their successors are independent of $\underline{C}_1 ... \underline{C}_M$ and their successors



From (2) and (3) it follows that

$$P(succ(\underline{Y}_1 .. \underline{Y}_N) \mid \underline{Y}_1 .. \underline{Y}_N) = \Pi\ P(succ(\underline{Y}_i) \mid \underline{Y}_i)$$

Bayesian Compositional Hierarchy (4)

$P(all)$ $= P(\underline{X} \mid succ(\underline{X}))\ P(succ(\underline{X}))$

$= P(\underline{X} \mid \underline{Y}_1 .. \underline{Y}_N)\ P(succ(\underline{X}))$ by Req 1

$= P(\underline{X} \mid \underline{Y}_1 .. \underline{Y}_N)\ P(\underline{Y}_1 .. \underline{Y}_N\ succ(\underline{Y}_1 .. \underline{Y}_N))$

$= P(\underline{X} \mid \underline{Y}_1 .. \underline{Y}_N)\ P(succ(\underline{Y}_1 .. \underline{Y}_N) \mid \underline{Y}_1 .. \underline{Y}_N)\ P(\underline{Y}_1 .. \underline{Y}_N)$

$= P(\underline{X} \mid \underline{Y}_1 .. \underline{Y}_N)\ \Pi\ P(succ(\underline{Y}_i) \mid \underline{Y}_i)\ P(\underline{Y}_1 .. \underline{Y}_N)$ by Req 2 + 3

➡ $P(succ(\underline{X}) \mid \underline{X}) = P(\underline{Y}_1 .. \underline{Y}_N \mid \underline{X})\ \Pi\ P(succ(\underline{Y}_i) \mid \underline{Y}_i)$

Recursive application gives:

$$P(\underline{Z}_0 .. \underline{Z}_M) = P(\underline{Z}_0) \prod_{i = 0 ... M} P(parts(\underline{Z}_i) \mid \underline{Z}_i)$$

Z_0 is a node and Z_i, $i = 1 .. M$ are its successors.

The complete JPD of an abstraction hierarchy can be computed from the conditional aggregate JPDs.

Probability changes may be propagated along tree-shaped hierarchy.

Alternative Formalization of Bayesian Compositional Hierarchy

External properties $\underline{Z}$ of an aggregate are determined by the functional mapping f: parts($\underline{Z}$) => $\underline{Z}$

⇒ P($\underline{Z}$ | parts($\underline{Z}$)) is known and fixed

The Bayesian Compositional Hierarchy factorization formula can be reformulated:

$$P(\underline{Z}_0 .. \underline{Z}_M) = \Pi \;\; P(\underline{Z}_i \mid parts(\underline{Z}_i)) \; C(parts(\underline{Z}_i))$$

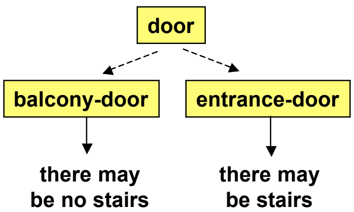
$$\text{where } C(\underline{Y}_1 .. \underline{Y}_N) = P(\underline{Y}_1 .. \underline{Y}_N) / \Pi \, P(\underline{Y}_i)$$

Given the probability distributions of the properties of individual parts, one can construct a hierarchy bottom-up by determining the correlations between parts belonging to an aggregate.

⇒ Unsupervised learning of aggregates

Choice of Alternative Specializations

Specializing a hypothesis

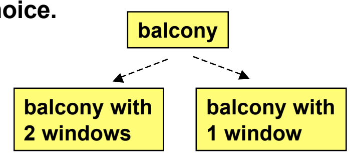


Disjunctive specializations can be modelled probabilistically, probability changes of one disjunctive branch may be propagated to the other branch.

Evidence assignment to one disjunctive branch forces specialization decision and must prohibit evidence assignment to the other branch.

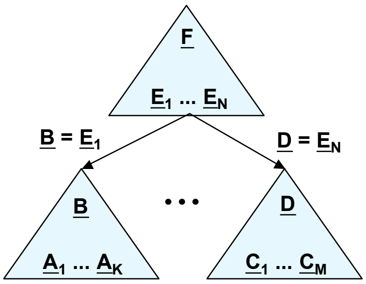
Currently, specialization decisions in SCENIC may be taken top-down, causing backtracking in the case of a wrong choice.

Aggregates with different cardinalities may be modelled as disjunctive specializations => the same applies.

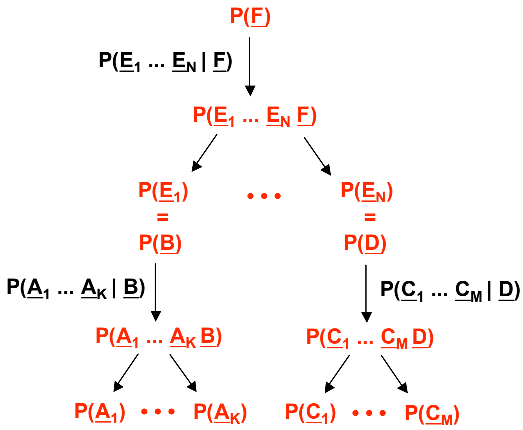


Top-down Initialization

Aggregate hierarchy subtree:



Sequence of computations:

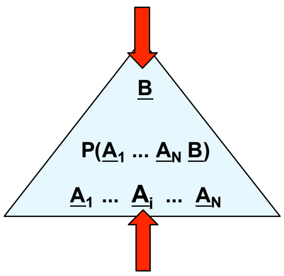


Change Propagation

After initialization, the state of each aggregate is represented by P(A_1 ... A_N) with marginalizations P(A_i), i = 1 ... N, and P(B).

A change has to be propagated if P(B) => P´(B) or P(A_i) => P´(A_i), some i.

Crisp evidence e for A_i is modelled as P(A_i = e) = 1 and P($A_i \neq$ e) = 0.

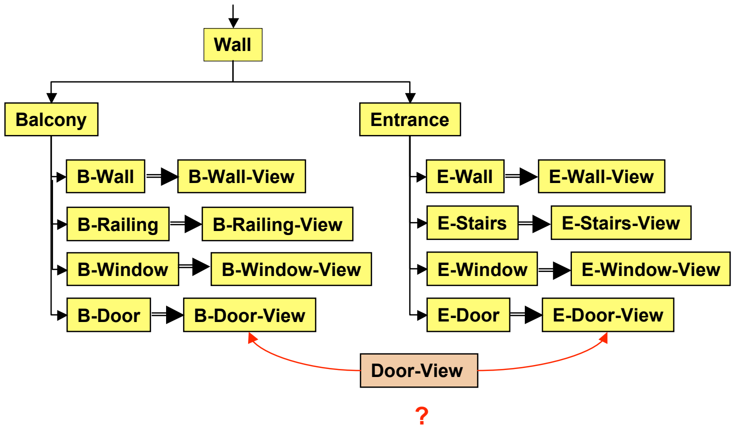


Propagating down:
P(B) => P´(B)
P´(A_1 ... A_N B) = P(A_1 ... A_N B) P´(B) / P(B)
followed by marginalizations

Propagating up:
P(A_i) => P´(A_i)
P´(A_1 ... A_N B) = P(A_1 ... A_N B) P´(A_i) / P(A_i)
followed by marginalizations

Preference Computation for Evidence Classification

- Probabilities within a branch may be compared without considering the rest of the compositional hierarchy
- Probabilities are updated after each decision and influence the following decisions



Best-first Evidence Classification

<u>Stepwise procedure</u>

A Choose evidence which allows most certain classification (reducing need for backtracking)

$\text{all } i \neq k\text{: } P(\text{view}_k \mid e) >> P(\text{view}_i \mid e)$

B If there is no probable classification for a given piece of evidence,
- perform backtracking to revise previous classifications, or
- request low-level validation of evidence

C Determine revised $P(\text{view}_i \mid e_j)$ after each classification
=> evidence propagation in probabilistic hierarchy

D Repeat steps A - D until task is completed
- evidence is exhausted
- scene interpretation is sufficiently certain
- specific interpretation request can be answered
- no conceptual model fits evidence

How to Determine Probability Distributions for Aggregates

1. Determine distributions for known crisp aggregates

   Two alternative approaches:

   a. Determine JPDs of internal and external properties by statistics (frequentist approach).

   b. Estimate JPDs based on human experiences and the mappings from internal to external properties.

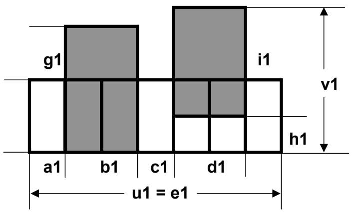
2. Learn aggregate concepts from scratch

   - Observe primitives, determine statistics
   - Build aggregate hierarchy by agglomerative clustering (use distance measure to establish Bayesian abstraction)
   - Derive higher-level probabilities from lower-level probabilities

Gaussian Aggregate Models

Uncertain aggregate properties can sometimes be roughly modelled as Gaussian densities.

Example:



Balcony probability densities:

$p_{b\text{-}door}(b1\ g1)$

$p_{b\text{-}window}(d1\ i1)$

$p_{railing}(b1\ g1)$

$p_{balcony\text{-}int}(a1\ b1\ c1\ d1\ e1\ f1\ g1\ h1\ i1)$

$p_{balcony\text{-}ext}(u1\ v1)$

$u1 = e1$
$v1 = h1 + i1$
} must be linear combination of parts properties

Probabilistic representation of the aggregate "balcony" by

$$P(a1\ b1\ c1\ d1\ e1\ f1\ g1\ h1\ i1 \mid u1\ v1)$$

parts properties (a1 b1 c1 d1 e1 f1 g1 h1 i1) — external aggregate properties (u1 v1)

Probabilistic Balcony Description

Specification of N(μ,Σ) for balcony properties by human estimates (unit = 1 dcm)

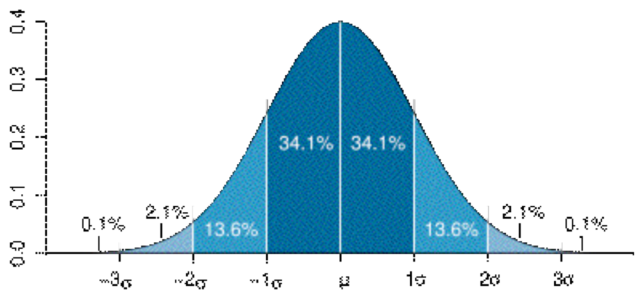
Means

a1	b1	c1	d1	e1	f1	g1	h1	i1	u1	v1
5	9	5	15	39	12	19	12	15	39	27

Covariances

	a1	b1	c1	d1	e1	f1	g1	h1	i1	u1	v1
a1	6,0	1,2	3,3	6,0	3,5	0,0	0,0	0,0	0,0	3,5	0,0
b1	1,2	2,3	1,2	5,3	2,1	0,0	0,4	0,0	1,2	2,1	1,2
c1	3,3	1,2	6,0	6,0	3,5	0,0	0,0	0,0	0,0	3,5	0,0
d1	6,0	5,3	6,0	60,0	11,0	0,0	0,0	0,0	8,5	11,0	8,5
e1	3,5	2,1	3,5	11,0	20,0	0,0	0,0	0,0	0,0	20,0	0,0
f1	0,0	0,0	0,0	0,0	0,0	1,0	0,0	0,0	0,0	0,0	0,0
g1	0,0	0,4	0,0	0,0	0,0	0,0	0,3	0,0	0,4	0,0	0,4
h1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	2,3	0,0	0,0	2,3
i1	0,0	1,2	0,0	8,5	0,0	0,0	0,4	0,0	6,0	0,0	6,0
u1	3,5	2,1	3,5	11,0	20,0	0,0	0,0	0,0	0,0	20,0	0,0
V1	0,0	1,2	0,0	8,5	0,0	0,0	0,4	2,3	6,0	0,0	8,3

Normal Distributions vs. Crisp Ranges



- **Gaussian range [-2σ .. 2σ] may be used for variables with crisp range type values**
- **Exploitation restricted to values in this range**

Propagation in Bayesian Compositional Hierarchies with Multivariate Gaussian Distributions

Multivariate Gaussian distribution: $P(\underline{G}) = N(\underline{\mu}_G, \Sigma_G)$

Remember general update formula for an aggregate with probability distribution $P(\underline{A}\ \underline{B})$ and change from $P(\underline{B})$ to $P'(\underline{B})$:

$$P'(\underline{A}\ \underline{B}) = P(\underline{A}\ \underline{B})\ P'(\underline{B})\ /\ P(\underline{B})$$

What is the new mean $\underline{\mu}_G'$ and covariance Σ_G' of a multivariate, if the distribution of a subset of the variables changes?

$$\Sigma_G = \begin{vmatrix} \Sigma_C & \Sigma_{CD} \\ \Sigma_{CD}^T & \Sigma_D \end{vmatrix} \quad \underline{\mu}_G = \begin{vmatrix} \underline{\mu}_C \\ \underline{\mu}_D \end{vmatrix}$$

Assume change $P(\underline{D}') = N(\underline{\mu}_D', \Sigma_D')$, then the new distribution is $N(\underline{\mu}_G', \Sigma_G')$

with

$$\underline{\mu}_C' = \underline{\mu}_C + \Sigma_{CD}\,\Sigma_D^{-1}\,(\underline{\mu}_D' - \underline{\mu}_D)$$

$$\Sigma_C' = \Sigma_C - \Sigma_{CD}\,\Sigma_D^{-1}\,\Sigma_{CD}^T + \Sigma_{CD}\,\Sigma_D^{-1}\,\Sigma_D'\,\Sigma_D^{-1}\,\Sigma_{CD}^T$$

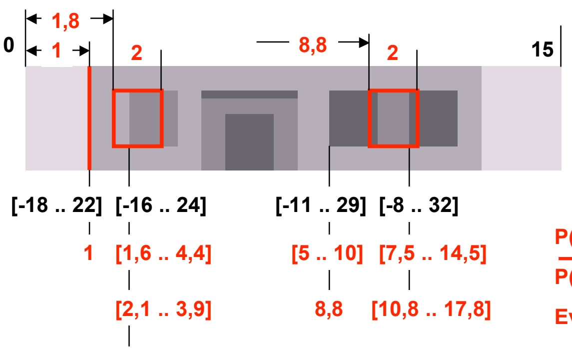
$$\Sigma_{CD}' = \Sigma_{CD}\,\Sigma_D^{-1}\,\Sigma_D'$$

=> Closed-form formulas allow efficient probability updates!

Example: Probability Propagation for Stepwise Facade Recognition

Interpretation process with probabilistic guidance using simple horizontal scene model

Value ranges correspond to [-2σ .. +2σ] of Gaussians



$$\frac{P(\text{Window2} \mid \text{Evidence2})}{P(\text{Window3} \mid \text{Evidence2})} = 1{,}36$$

Evidence3 is incompatible with model!