Qualitative Probabilistic Matching with Hierarchical Descriptions

Clinton Smyth and David Poole

Georeference Online and UBC

http://www.georeferenceonline.com/

☞

Overview

- Levels of abstraction and detail
- Qualitative probabilities
- Instances and Models
- Matching:
  - finding most likely models for an instance
  - finding most likely instances for a model

Motivating Example: minerals exploration

- In geology, different parts of the world are described at various levels of abstraction and detail.
- People spend careers developing models of where certain minerals can be found.
- We want to determine which described instances match various models.
- It is people that must make decisions; computers help by narrowing down the search space and explaining and justifying the potential matches.

The Problem

Given

- a set of instances described at various levels of abstraction and details
- a set of qualitative-probabilistic models at various levels of abstraction and detail

find

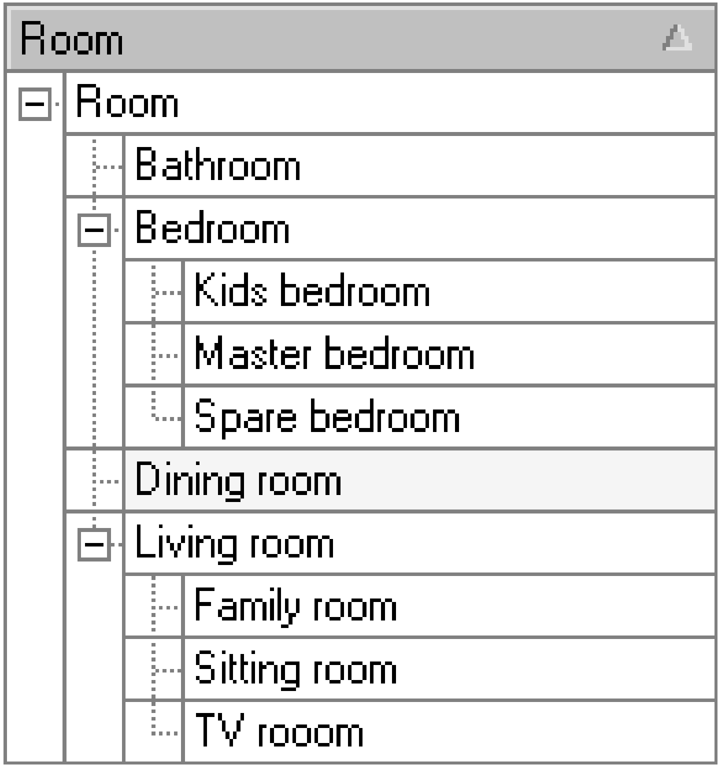
- the best model(s) that match an instance
- the best instance(s) that match a model

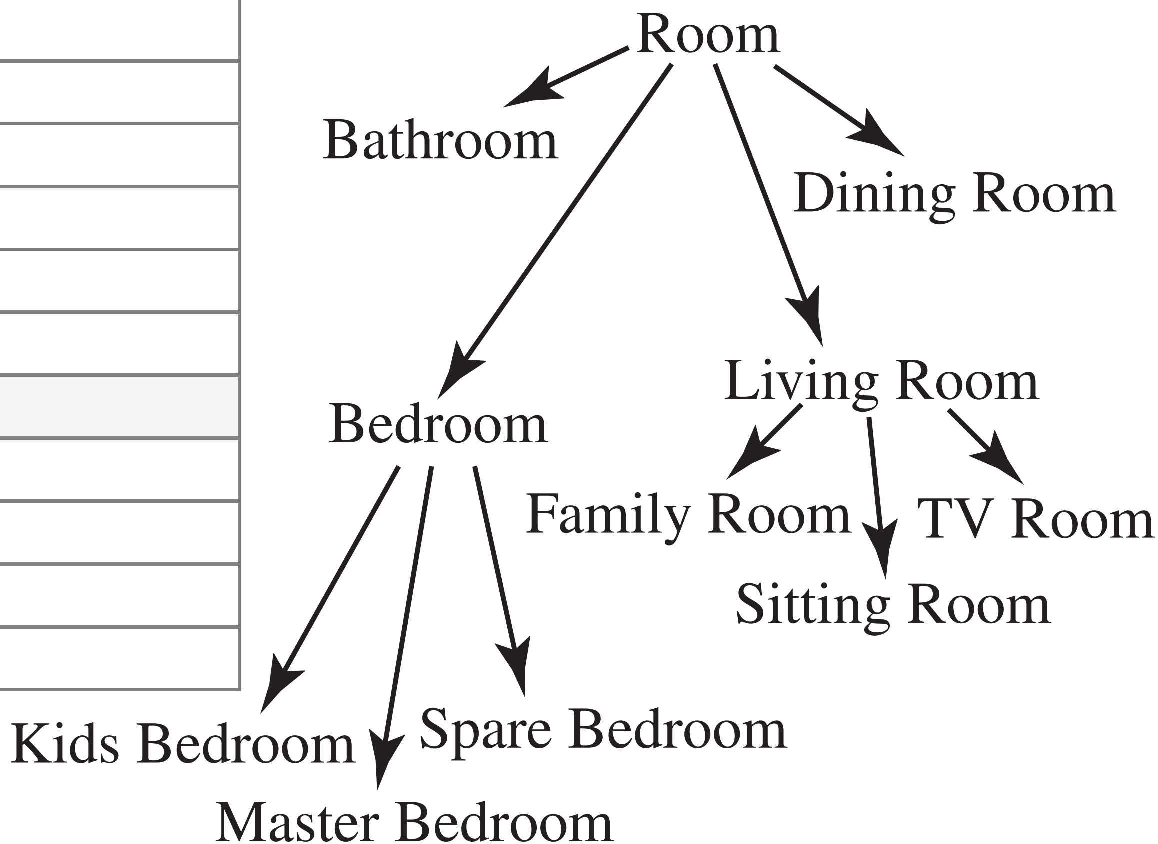
and explain the match to enable people to make informed decisions.

Levels of abstraction

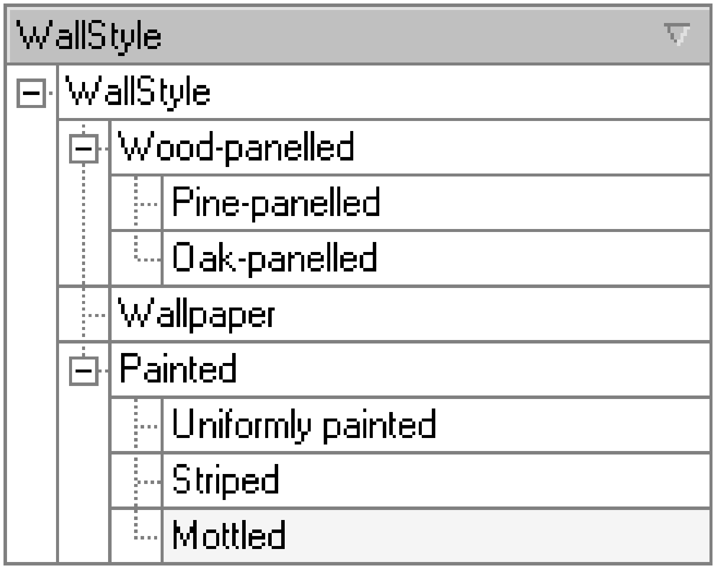
- Different parts of the world are described at different levels of abstraction $\rightarrow$ taxonomic or value hierarchy.
- We assume an ontology that includes taxonomic hierarchies that specify the vocabulary for different levels of abstraction.

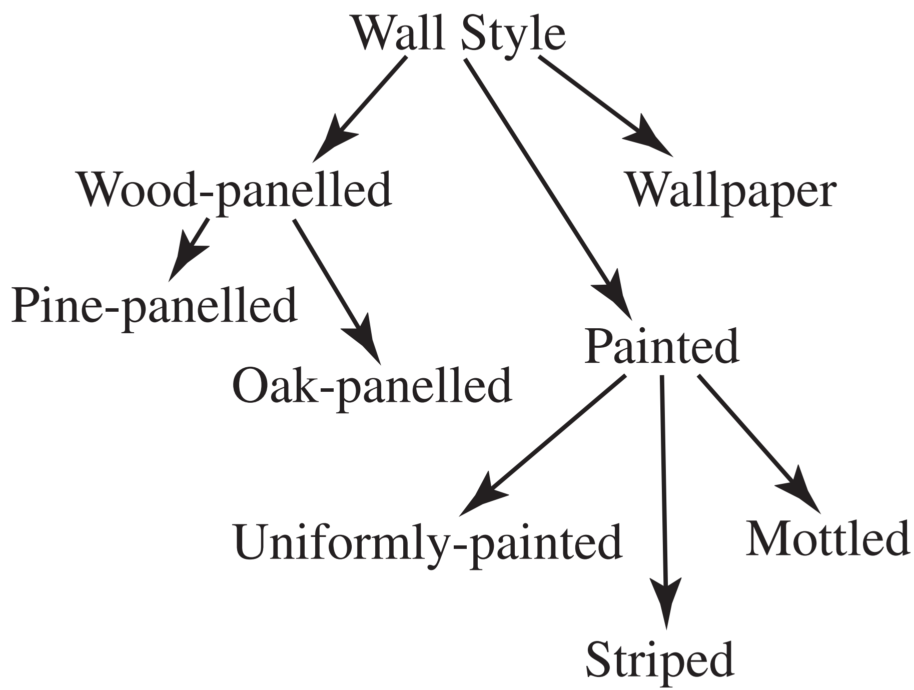
Example: Room abstraction hierarchy





Example: WallStyle abstraction hierarchy





Level of Detail

- The same thing can be described at various levels of detail:
  - An apartment that contains a bath.
  - An apartment that contains a bathroom that contains a bath.
  - An apartment that contains a mottled master bedroom and a uniformly-painted bathroom that contains a green bath.

Qualitative Probabilities

- The role of models is to help make decisions.
- Decisions are based on probabilities and utilities.
- Decisions are made by people, based on their own subjective probabilities and utilities.
- People are overwhelmed by the combinatorics; the number of instances and models and potential matches.
- We provide an approximation to the probabilities, enumerate the most likely approximate matches and justify our conclusions, to help people make decisions.

Kappa Calculus

- The kappa calculus is a qualitative approximation to probabilistic reasoning.
- It can be seen a order-of-magnitude probabilities.
  - Multiplication in probability corresponds to addition in the kappa calculus.
  - Addition in probability corresponds to maximisation in the kappa calculus.
- When using the kappa calculus to compare models or compare instances, the scale and the zero are arbitrary.

Qualitative Probabilities

- The models are specified using a 5 point scale to specify the qualitative probability of a proposition:
  - **always** you are very surprised if it's false.
  - **usually** you are somewhat surprised if it's false.
  - **sometimes** you aren't surprised if it's true or false
  - **rarely** you are somewhat surprised if it's true.
  - **never** you are very surprised if it's true.
- The output from the match is a numerical value.

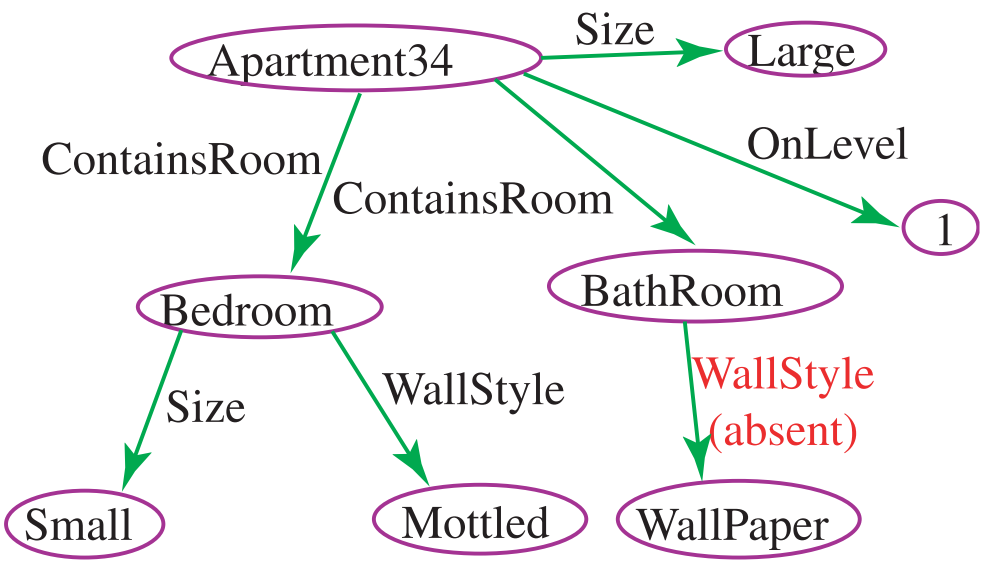
Model and Instance Representations

- Models and instances are described as

$$\langle \mathit{Object}, \mathit{Attribute}, \mathit{Value}, \mathit{Frequency}, \mathit{Citation} \rangle$$

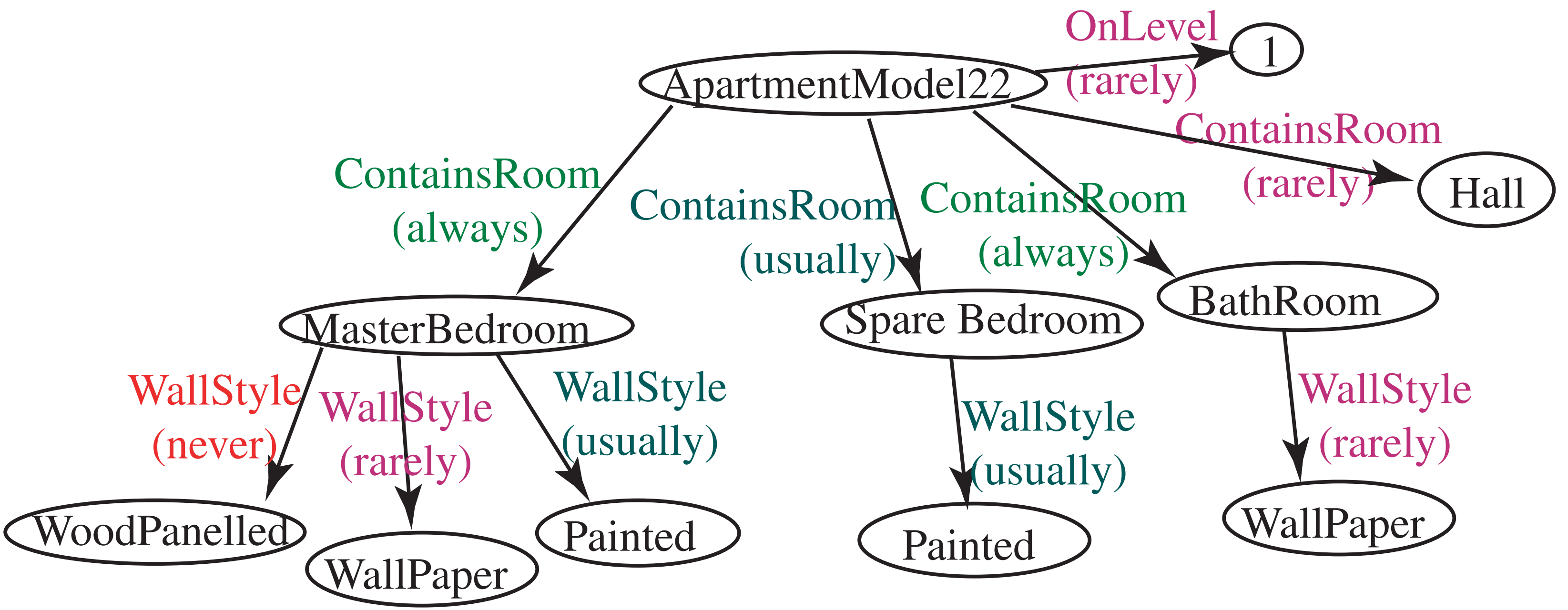
- The frequency is either:
  - truth value for instances
  - qualitative probability for models

Example Instance



Apartment34's Attribute - Value	Present or Absent
ContainsRoom - Bedroom	present
Size - Small	present
WallStyle - Mottled	present
ContainsRoom - BathRoom	present
WallStyle - WallPaper	absent
Size - Large	present
OnLevel - 1	present

Example Model



ApartmentModel22's Attribute - Value	Expected Frequency
ContainsRoom - MasterBedroom	always
WallStyle - WoodPanelled	never
WallStyle - WallPaper	rarely
WallStyle - Painted	usually
ContainsRoom - SpareBedroom	usually
WallStyle - Painted	usually
ContainsRoom - BathRoom	always
WallStyle - WallPaper	rarely
ContainsRoom - Hall	rarely
OnLevel - 1	rarely

Inference

- We want to find $P(instance|model)$ - or its qualitative counterpart.
- We can use Bayes' theorem to compute:

$$P(model|instance) \propto P(instance|model)P(model)$$

- Usually we want the most likely instance(s) given the model or the most likely model(s) given the instance.

Inference

- If not for
  - different levels of abstraction,
  - different levels of detail (missing attribute values, multiple values for an attribute),

  the qualitative probability of an instance given a model is the sum of the "surprises" of the instance with respect to the qualitative probabilities.
- The "zero" is defined as the value of the match of the empty description with the empty model.

Descriptions and Instances

Suppose we have instances:

- i_1: something that is painted
- i_2: a painted room,
- i_3: a painted bathroom,
- i_4: a mottled bathroom

$P(i_1|E) \geq P(i_2|E) \geq P(i_3|E) \geq P(i_4|E)$ for any E.

- What if the model says we (always) want a painted bathroom?

Abstacting Instances and Models

- If an instance has a description at a high level of abstraction, it means we don't know which subclass it is.
- If a model has a description at a high level of abstraction, it means we don't care which subclass it is.
- For the probability of the instance, we want the probability of the actual instance, not the probability of a description.

Meaning of parts

We make the assumption that

- In a single description, different descriptions of parts pertain to different parts.
- We do not assume that all parts are given.
- Example: a description of an instance that contains a mottled room and a painted bathroom.
  It actually contains (at least) two rooms
  (as opposed to containing one room that is a mottled bathroom).

Matching Parts: Example

- Model: Joe always likes an apartment that contains a mottled bathroom.
- Instance: an apartment with a mottled room and a painted bathroom.

Hypotheses:

- the mottled room is a bathroom,
- the painted bathroom is mottled,
- there is another room that is a mottled bathroom
- there is no mottled bathroom

Example match

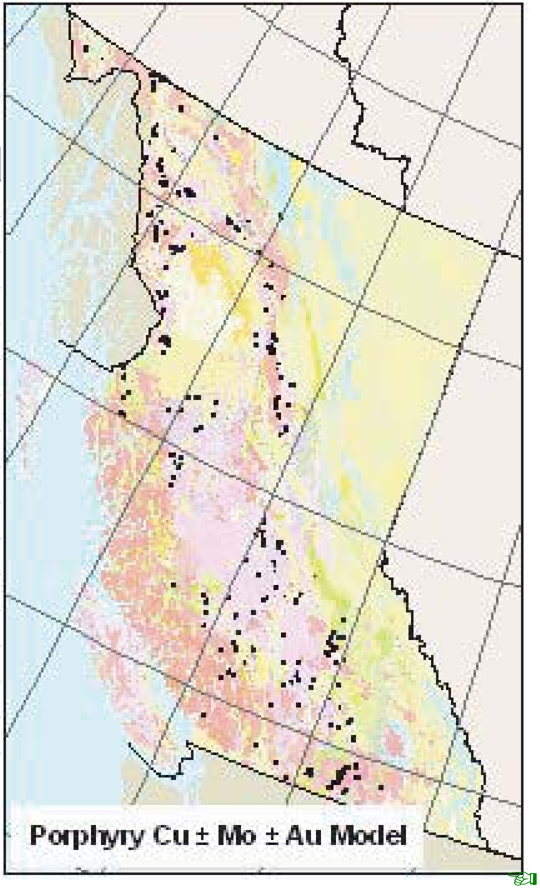
ApartmentModel22: Attribute	ApartmentModel22's Value	ApartmentModel22: Expected Frequency	Apartment34: Attribute	Apartment34's Value	Apartment34: Present or Absent	Match Type
ContainsRoom	SpareBedroom	usually				
WallStyle	Painted	usually				
ContainsRoom	BathRoom	always	ContainsRoom	BathRoom	present	exact
WallStyle	WallPaper	rarely				
ContainsRoom	Hall	rarely				
ContainsRoom	MasterBedroom	always	ContainsRoom	Bedroom	present	maybeAKO
			Size	Small	present	
WallStyle	WallPaper	rarely				
WallStyle	Painted	usually	WallStyle	Mottled	present	exactAKO
WallStyle	WoodPanelled	never				
OnLevel	1	rarely	OnLevel	1	present	conflict
			Size	Large	present	

Example match: 1 model, 5000 Instances

GEOREFERENCE Online Ltd

Rankings for Porphyry Cu ± Mo ± Au Model Against 5000 Instances

Match Object	Rank	Overall Score	Penalties	Rewards
Cluster 0	1	1.5	144	29500
Cluster 470	2	1.5	150	29500
Cluster 948	3	1.5	161	29500
Cluster 2266	4	1.5	144	29167
Cluster 2214	5	1.5	142	28833
Cluster 3387	6	1.4	142	27500
Cluster 3141	7	1.4	143	27500
Cluster 477	8	1.3	244	25167
Cluster 554	9	1.3	152	25000
Cluster 1461	10	1.3	175	25000
...				
...				
Cluster 2075	247	0.8	162	14000
Cluster 3389	248	0.8	163	14000
Cluster 2153	249	0.8	164	14000
Cluster 3004	250	0.8	166	14000



Conclusions

- This paper has outlined issues when doing (qualitative) probabilistic inference with a rich hypothesis space.
- Issues: multiple levels of abstraction and levels of detail.
- This is a working system in minerals explorations with thousands of instances and hundreds of models. Also used for landslides and mapping....
- Mix ontologies and probabilistic reasoning.