Semantic Science: ontologies, data and probabilistic theories

David Poole

Department of Computer Science,
University of British Columbia

January 2010

For when I am presented with a false theorem, I do not need to examine or even to know the demonstration, since I shall discover its falsity *a posteriori* by means of an easy experiment, that is, by a calculation, costing no more than paper and ink, which will show the error no matter how small it is. . .

And if someone would doubt my results, I should say to him: "Let us calculate, Sir," and thus by taking to pen and ink, we should soon settle the question.

—Gottfried Wilhelm Leibniz [1677]

Example: medical diagnosis

Example: people give symptoms and want to know what is wrong with them.

Current Practice (Google)	Semantic Science Alternative
— describe symptoms using keywords — results ranked by popularity (pagerank) and usually appeal to authority — text results	

Example: medical diagnosis

Example: people give symptoms and want to know what is wrong with them.

Current Practice (Google)	Semantic Science Alternative
— describe symptoms using keywords	— use ontologies
— results ranked by popularity (pagerank) and usually appeal to authority	— predictions ranked by relevance and fit to data
— text results	— probabilistic predictions with references to sources

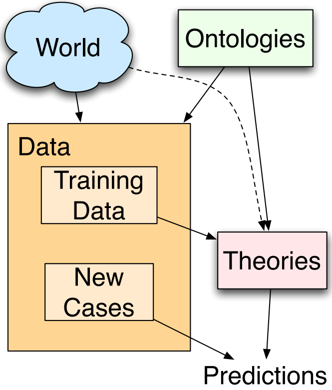
Outline

1. Semantic Science Overview
    - Ontologies
    - Data
    - Theories
2. Representing Probabilistic Theories
    - Feature-based Theories
    - First-order probabilistic models
    - Probabilities with Ontologies
    - Existence and Identity Uncertainty
3. Pragmatics of Real Theories

Notational Minefield

- Theory / hypothesis / model / law (Science)
- Variable (probability and logic and programming languages)
- Model (science, probability and logic)
- Parameter (mathematics and statistics)
- Domain (science and logic and probability and mathematics)
- Object/class (object-oriented programming and ontologies)
- $=$ (probability and logic)
- First-order (logic and dynamical systems)

Semantic Science



- Ontologies represent the meaning of symbols.
- Data that adheres to an ontology is published.
- Theories that make (probabilistic) predictions on data are published.
- Data can be used to evaluate theories.
- Theories make predictions on new cases.
- All evolve in time.

AI Traditions

- Expert Systems of the 70's and 80's
  - Probabilistic models and machine learning.
    Bayesian networks, Bayesian X...
  - Ontologies and Knowledge Representations.
    Description logic, X logic...

AI Traditions

- Expert Systems of the 70's and 80's
  - Probabilistic models and machine learning.
    Bayesian networks, Bayesian X...
  - Ontologies and Knowledge Representations.
    Description logic, X logic...
- Machine Learning
  - Heterogeneous data sets with rich ontologies
  - Persistent theories built by humans and automatically

Science in Broadest Sense

I mean *science* in the broadest sense:

- where and when landslides occur
- where to find gold
- what errors students make
- disease symptoms, prognosis and treatment
- what companies will be good to invest in
- what apartment Mary would like

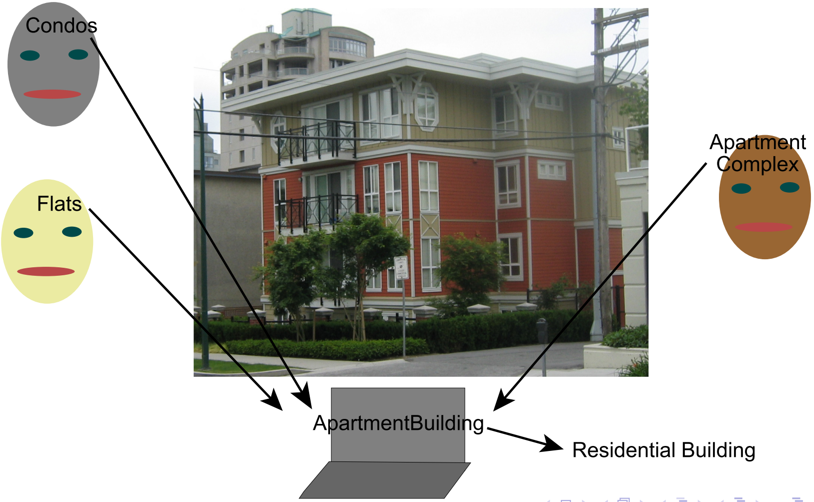
Outline

1. Semantic Science Overview
   - Ontologies
   - Data
   - Theories
2. Representing Probabilistic Theories
   - Feature-based Theories
   - First-order probabilistic models
   - Probabilities with Ontologies
   - Existence and Identity Uncertainty
3. Pragmatics of Real Theories

Ontologies

- In philosophy, ontology the study of existence.
- In CS, an ontology is a (formal) specification of the meaning of the vocabulary used in an information system.
- Ontologies are needed so that information sources can inter-operate at a semantic level.

Ontologies


Condos
Flats
Apartment Complex
ApartmentBuilding
Residential Building

Choosing Individuals and Relations in Logic

First-order logical languages allow many different ways of representing facts.

E.g., How to represent: "Pen #7 is red."

Choosing Individuals and Relations in Logic

First-order logical languages allow many different ways of representing facts.

E.g., How to represent: "Pen #7 is red."

- $red(pen_7)$. It's easy to ask "What's red?"
  Can't ask "what is the color of pen_7?"

Choosing Individuals and Relations in Logic

First-order logical languages allow many different ways of representing facts.

E.g., How to represent: "Pen #7 is red."

- *red*(pen_7). It's easy to ask "What's red?"
  Can't ask "what is the color of pen_7?"
- *color*(pen_7, *red*). It's easy to ask "What's red?"
  It's easy to ask "What is the color of pen_7?"
  Can't ask "What property of pen_7 has value *red*?"

Choosing Individuals and Relations in Logic

First-order logical languages allow many different ways of representing facts.

E.g., How to represent: "Pen #7 is red."

- $red(pen_7)$. It's easy to ask "What's red?"
  Can't ask "what is the color of pen_7?"
- $color(pen_7, red)$. It's easy to ask "What's red?"
  It's easy to ask "What is the color of pen_7?"
  Can't ask "What property of pen_7 has value red?"
- $prop(pen_7, color, red)$. It's easy to ask all these questions.

Choosing Individuals and Relations in Logic

First-order logical languages allow many different ways of representing facts.

E.g., How to represent: "Pen #7 is red."

- $red(pen_7)$. It's easy to ask "What's red?"
  Can't ask "what is the color of pen_7?"
- $color(pen_7, red)$. It's easy to ask "What's red?"
  It's easy to ask "What is the color of pen_7?"
  Can't ask "What property of pen_7 has value red?"
- $prop(pen_7, color, red)$. It's easy to ask all these questions.

$prop(Individual, Property, Value)$ is the only relation needed:
$\langle Individual, Property, Value \rangle$ triples, Semantic network, entity relationship model, ...

Reification

- To represent $scheduled(cs422, 2, 1030, cc208)$. "section 2 of course $cs422$ is scheduled at 10:30 in room $cc208$."
- Let $b123$ name the booking:

$$prop(b123, course, cs422).$$
$$prop(b123, section, 2).$$
$$prop(b123, time, 1030).$$
$$prop(b123, room, cc208).$$

- We have reified the booking.
- Reify means: to make into an individual.

Semantic Web Ontology Languages

- RDF — language for triples in XML. Everything is a resource (with URI)
- RDF Schema — define resources in terms of each other: class, type, subClassOf, subPropertyOf, collections. . .
- OWL — allows for equality statements, restricting domains and ranges of properties, transitivity, cardinality. . .
- OWL-Lite, OWL-DL, OWL-Full

Main Components of an Ontology

- Individuals: the objects in the world (not usually specified as part of the ontology)
- Classes: sets of (potential) individuals
- Properties: between individuals and their values

Aristotelian definitions

Aristotle [350 B.C.] suggested the definition if a class C in terms of:

- Genus: the super-class
- Differentia: the attributes that make members of the class C different from other members of the super-class

"If genera are different and co-ordinate, their differentiae are themselves different in kind. Take as an instance the genus 'animal' and the genus 'knowledge'. 'With feet', 'two-footed', 'winged', 'aquatic', are differentiae of 'animal'; the species of knowledge are not distinguished by the same differentiae. One species of knowledge does not differ from another in being 'two-footed'."

Aristotle, *Categories*, 350 B.C.

An Aristotelian definition

- An **apartment building** is a **residential building** with **multiple units** and **units are rented**.

$$\begin{aligned} ApartmentBuilding &\equiv ResidentialBuilding\& \\ & NumUnits = many\& \\ & Ownership = rental \end{aligned}$$

  NumUnits is a property with domain *ResidentialBuilding* and range $\{one, two, many\}$
  Ownership is a property with domain *Building* and range $\{owned, rental, coop\}$.

- All classes can be defined in terms of properties.

Outline

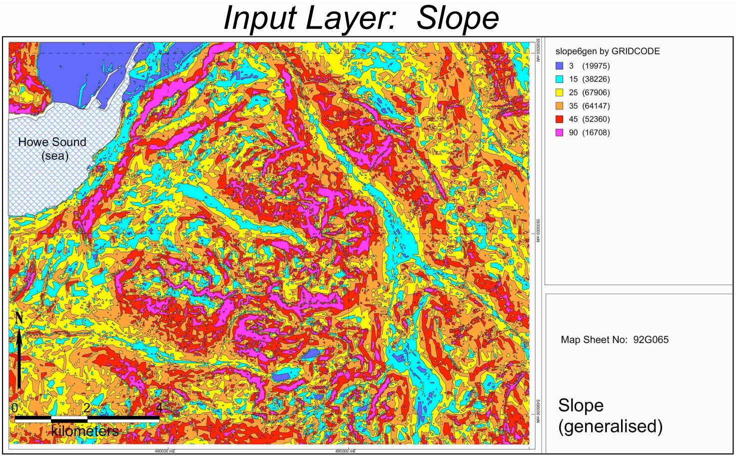
1. Semantic Science Overview
   - Ontologies
   - Data
   - Theories
2. Representing Probabilistic Theories
   - Feature-based Theories
   - First-order probabilistic models
   - Probabilities with Ontologies
   - Existence and Identity Uncertainty
3. Pragmatics of Real Theories

Data

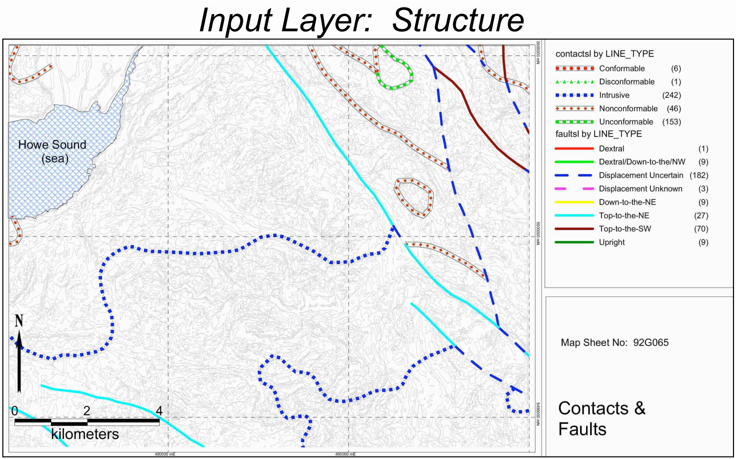
Real data is messy!

- Multiple levels of abstraction
- Multiple levels of detail
- Uses the vocabulary from many ontologies: rocks, minerals, top-level ontology,. . .
- Rich meta-data:
  - Who collected each datum? (identity and credentials)
  - Who transcribed the information?
  - What was the protocol used to collect the data? (Chosen at random or chosen because interesting?)
  - What were the controls — what was manipulated, when?
  - What sensors were used? What is their reliability and operating range?

Example Data, Geology



Example Data, Geology



http://www.vsto.org/

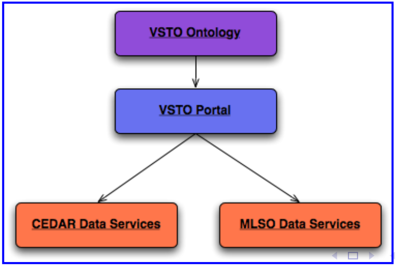
Welcome to the Virtual Solar Terrestrial Observatory

The Virtual Solar Terrestrial Observatory (VSTO) is a unified semantic environment serving data from diverse data archives in the fields of solar, solar-terrestrial, and space physics (SSTSP), currently:

- Upper atmosphere data from the **CEDAR** (Coupling, Energetics and Dynamics of Atmospheric Regions) archive
- Solar corona data from the **MLSO** (Mauna Loa Solar Observatory) archive

The VSTO portal uses an underlying ontology (i.e. an organized knowledge base of the SSTSP domain) to present a general interface that allows selection and retrieval of products (ascii and binary data files, images, plots) from heterogenous external data services.

VSTO Data Access



Data is theory-laden

- Sapir-Whorf Hypothesis [Sapir 1929, Whorf 1940]: people's perception and thought are determined by what can be described in their language. (Controversial in linguistics!)
- A stronger version for information systems:

  What is stored and communicated by an information system is constrained by the representation and the ontology used by the information system.

- Ontologies must come logically prior to the data.
- Data can't make distinctions that can't be expressed in the ontology.
- Different ontologies result in different data.

Outline

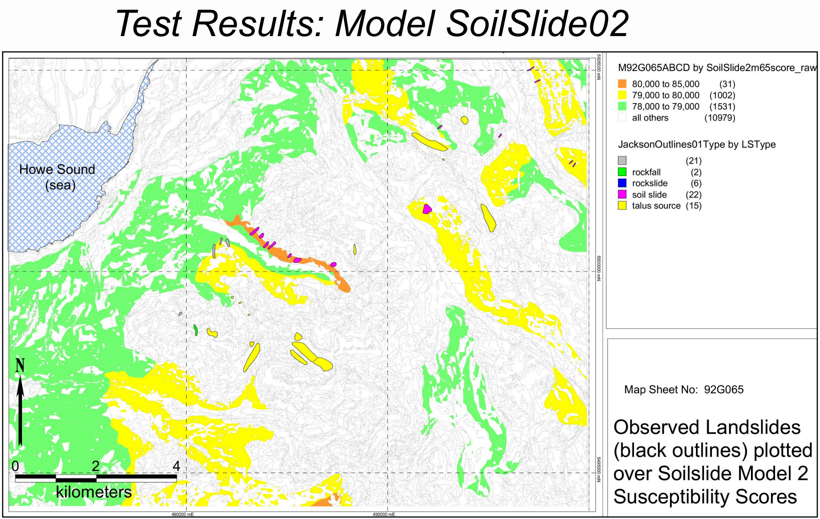
1. Semantic Science Overview
   - Ontologies
   - Data
   - Theories
2. Representing Probabilistic Theories
   - Feature-based Theories
   - First-order probabilistic models
   - Probabilities with Ontologies
   - Existence and Identity Uncertainty
3. Pragmatics of Real Theories

Theories make predictions on data

Hypotheses and theories are procedures that make prediction on data.

- Theories can make various predictions about data:
  - definitive predictions
  - point probabilities
  - probability ranges
  - ranges with confidence intervals
  - qualitative predictions
- For each prediction type, we need ways to judge predictions on data
- Users can use whatever criteria they like to evaluate theories (e.g., taking into account simplicity and elegance)
- We want to input hypotheses and output theories.

Example Prediction from a Theory



Theory Ensembles

- How can we compare theories that differ in their generality?
- Theory A makes predictions about all cancers. Theory B makes predictions about lung cancers. Should the comparison between A and B take into account A's predictions on non-lung cancer?

Theory Ensembles

- How can we compare theories that differ in their generality?
- Theory A makes predictions about all cancers.
  Theory B makes predictions about lung cancers.
  Should the comparison between A and B take into account A's predictions on non-lung cancer?
- What about theory C: *if lung cancer, use B's prediction, else use A's prediction*?

Theory Ensembles

- How can we compare theories that differ in their generality?
- Theory A makes predictions about all cancers.
  Theory B makes predictions about lung cancers.
  Should the comparison between A and B take into account A's predictions on non-lung cancer?
- What about theory C: *if lung cancer, use B's prediction, else use A's prediction*?
- Proposal: make theory ensembles the norm.
  - Judge theories by how well they fit into ensembles.
  - Ensembles can be judged by simplicity.
  - Theory designers don't need to game the system by manipulating the generality of theories

Dynamics of Semantic Science

- Anyone can design their own ontologies.
  — People vote with their feet what ontology they use.
  — Need for semantic interoperability leads to ontologies with mappings between them.
- Ontologies evolve with theories:
  A theory hypothesizes unobserved features or useful distinctions
  $\longrightarrow$ add these to an ontology
  $\longrightarrow$ other researchers can refer to them
  $\longrightarrow$ reinterpretation of data
- Ontologies can be judged by the predictions of the theories that use them
  — the role of the vocabulary is to describe useful distinctions

Levels of Semantic Science

0. Deterministic semantic science where all of the theories make definitive predictions.
1. Feature-based semantic science, with non-deterministic predictions about feature values of data.
2. Relational semantic science, with predictions about the properties of objects and relationships among objects.
3. First-order semantic science, with predictions about the existence of objects, universally quantified statements and relations.

Outline

1. Semantic Science Overview
   - Ontologies
   - Data
   - Theories
2. Representing Probabilistic Theories
   - Feature-based Theories
   - First-order probabilistic models
   - Probabilities with Ontologies
   - Existence and Identity Uncertainty
3. Pragmatics of Real Theories

Probabilistic Prediction

- The role of models in prediction: Given a description of a new case,

$$
\begin{aligned}
& P(prediction|description) \\
= & \sum_{m \in Models} \left(\begin{array}{l} P(prediction|m\&description)\times \\ P(m|description) \end{array} \right)
\end{aligned}
$$

Models is a set of mutually exclusive and covering set of hypotheses.

Probabilistic Prediction

- The role of models in prediction: Given a description of a new case,

$$
\begin{aligned}
& P(prediction \mid description) \\
= & \sum_{m \in Models} \left(\begin{array}{l} P(prediction \mid m \& description) \times \\ P(m \mid description) \end{array} \right)
\end{aligned}
$$

  Models is a set of mutually exclusive and covering set of hypotheses.
- What features of the description are predictive?
- How do the features interact?
- What are the appropriate probabilities? (How can these be learned with limited data?)

Representing Uncertainty: Bayesian belief networks

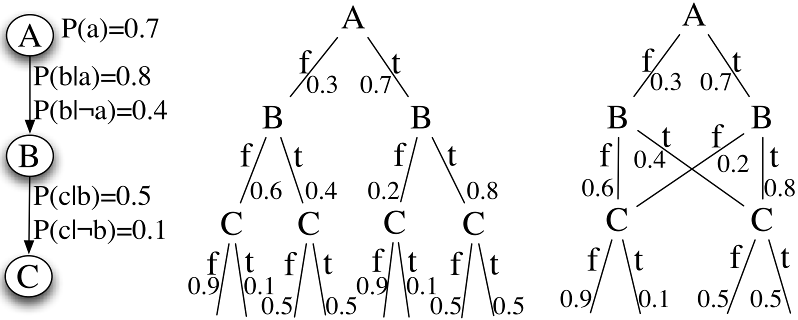
What:

- A belief network is a graphical representation of dependence amongst a set of random variables.

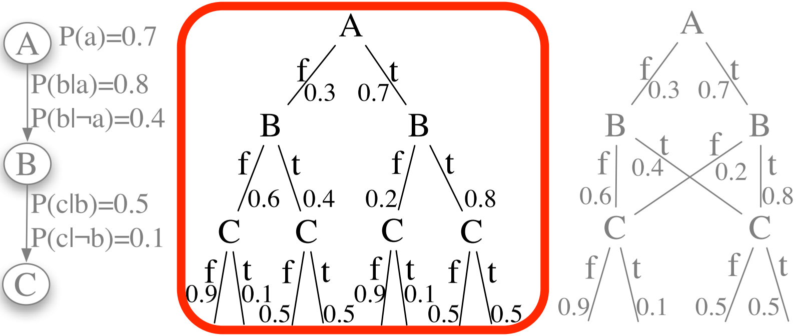
Why:

- Often the natural representation: independence represents causal structure
- Probabilities can be understood and learned locally
- We can exploit the structure for efficient inference

Semantic Tree



Semantic Tree



↑

semantic tree

event tree

decision tree. . .

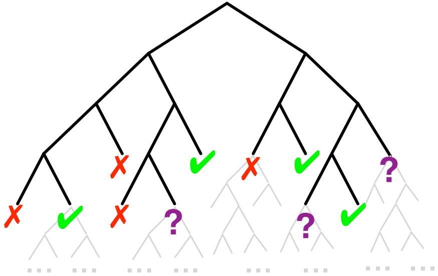
Semantic tree

- Nodes are propositions or discrete variables
- Child for each value in domain
- There is a probability distribution over the children of each node
- Each finite path from the root corresponds to a formula
- Each finite path from the root has a probability that is the product of the probabilities in the path

A generative model generates a semantic tree.

Infinite Semantic Tree

Given a proposition α:



✔ *path* $\models \alpha$
✗ *path* $\models \neg\alpha$
? otherwise

The probability of α is well defined if
for all $\epsilon > 0$
there is a finite sub-tree that can answer α in
$> 1 - \epsilon$ of the probability mass.

Outline

1. Semantic Science Overview
   - Ontologies
   - Data
   - Theories
2. Representing Probabilistic Theories
   - Feature-based Theories
   - First-order probabilistic models
   - Probabilities with Ontologies
   - Existence and Identity Uncertainty
3. Pragmatics of Real Theories

Data

Data is of observations of a world.
Meta-data about observations includes:

- The context in which the data was collected.
- The features that this data makes predictions about (the dependent variables).
- The features that were controlled for in the data (the independent variables).

Theories

A theory makes predictions about some feature values.
A theory includes:

- A context in which specifies when it can be applied.
- A set of input features about which it does not make predictions.
- A set of output features to predict (as a function of the input features).

Example

Consider the following theories:

- T_1 predicts the prognosis of people with lung cancer.
- T_2 predicts the prognosis of people with cancer.
- T_3 is the null hypothesis that predicts the prognosis of people in general.
- T_4 predicts (probabilistically) whether people with cancer have lung cancer, as a function of coughing.
- T_5 predicts (probabilistically) whether people have cancer.

What should be used to predict the prognosis of a patient with observed symptoms?

Theory Ensembles

To make a prediction, multiple theories need to be used together—theory ensemble.

A theory ensemble T needs to satisfy the following properties:

- T is coherent: it does not rely on the value of a feature in a context where the features is not defined
- T is consistent: it does not make different predictions for any feature in any context.
- T is predictive: it makes a prediction in every context that is possible.
- T is minimal.

Prototype Feature-based Ensemble

- An (oversimplified) definition of theory ensembles is a set of $\langle c, t\rangle$ pairs, where t is a theory and c is a proposition.
- Pair $\langle c, t\rangle$ means theory t is used to predict in context c.

Example

- T_1 predicts the prognosis of people with lung cancer.
- T_2 predicts the prognosis of people with cancer.
- T_3 is the null hypothesis that predicts the prognosis of people in general.
- T_4 predicts (probabilistically) whether people with cancer have lung cancer, as a function of coughing.
- T_5 predicts (probabilistically) whether people have cancer.

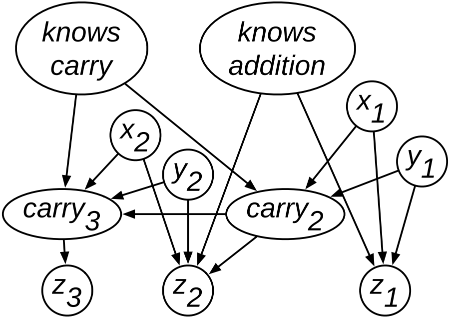
A possible theory ensemble is:
$\{\langle person, T_5\rangle, \langle \neg has\_cancer, T_3\rangle, \langle has\_cancer, T_4\rangle, \langle has\_lung\_cancer, T_1\rangle, \langle has\_cancer \wedge \neg has\_lung\_cancer, T2\rangle\}$

Outline

1. Semantic Science Overview
    - Ontologies
    - Data
    - Theories
2. Representing Probabilistic Theories
    - Feature-based Theories
    - First-order probabilistic models
    - Probabilities with Ontologies
    - Existence and Identity Uncertainty
3. Pragmatics of Real Theories

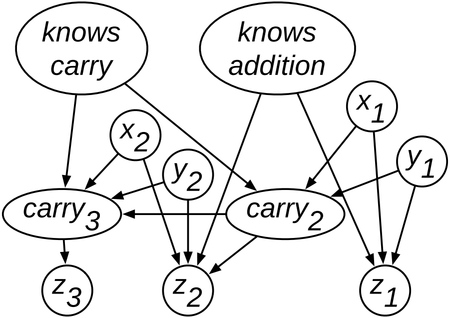
Predicting students errors

$$\begin{array}{cccc} & & x_2 & x_1 \\ + & & y_2 & y_1 \\ \hline & z_3 & z_2 & z_1 \end{array}$$



Predicting students errors

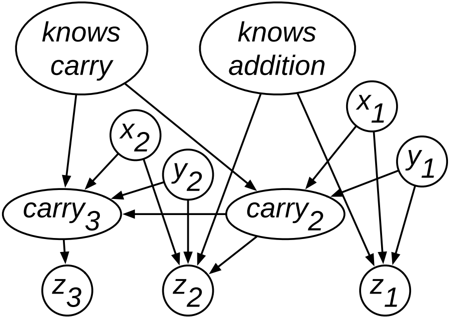
$$
\begin{array}{rccc}
 & & x_2 & x_1 \\
+ & & y_2 & y_1 \\
\hline
 & z_3 & z_2 & z_1
\end{array}
$$



What if there were multiple digits

Predicting students errors

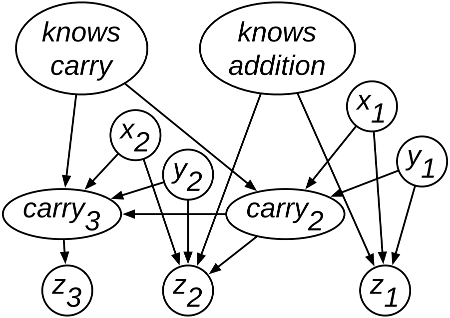
$$\begin{array}{rrrr} & & x_2 & x_1 \\ + & & y_2 & y_1 \\ \hline & z_3 & z_2 & z_1 \end{array}$$



What if there were multiple digits, problems

Predicting students errors

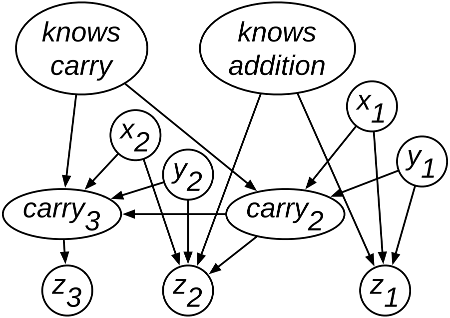
$$\begin{array}{rrrr} & & x_2 & x_1 \\ + & & y_2 & y_1 \\ \hline & z_3 & z_2 & z_1 \end{array}$$



What if there were multiple digits, problems, students

Predicting students errors

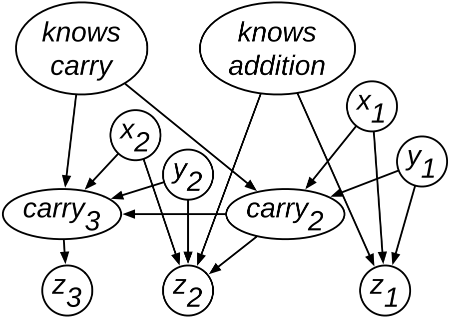
$$\begin{array}{rrrr} & & x_2 & x_1 \\ + & & y_2 & y_1 \\ \hline & z_3 & z_2 & z_1 \end{array}$$



What if there were multiple digits, problems, students, times?

Predicting students errors

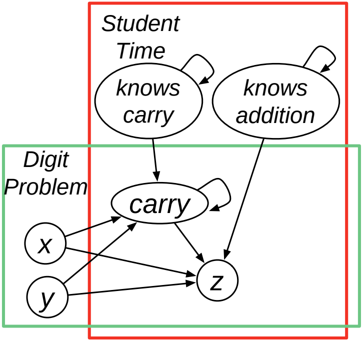
$$\begin{array}{rrrr} & & x_2 & x_1 \\ + & & y_2 & y_1 \\ \hline & z_3 & z_2 & z_1 \end{array}$$



What if there were multiple digits, problems, students, times?
How can we build a model before we know the individuals?

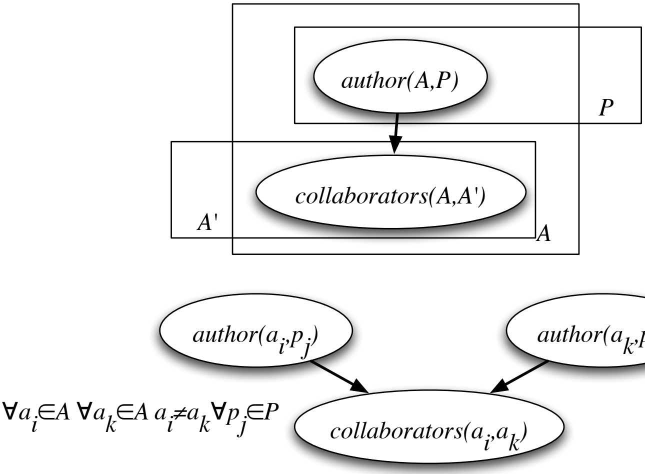
Multi-digit addition with parametrized BNs / plates

$$\begin{array}{rcccc} & x_{j_x} & \cdots & x_2 & x_1 \\ + & y_{j_z} & \cdots & y_2 & y_1 \\ \hline & z_{j_z} & \cdots & z_2 & z_1 \end{array}$$



Random Variables: $x(D, P)$, $y(D, P)$, $knowsCarry(S, T)$, $knowsAddition(S, T)$, $carry(D, P, S, T)$, $z(D, P, S, T)$ for each: digit D, problem P, student S, time T

Creating Dependencies: Relational Structure



Independent Choice Logic

- A language for first-order probabilistic models.
- Idea: combine logic and probability, where all uncertainty in handled in terms of Bayesian decision theory, and logic specifies consequences of choices.
- History: parametrized Bayesian networks, abduction and default reasoning $\longrightarrow$ probabilistic Horn abduction (IJCAI-91); richer language (negation as failure + choices by other agents $\longrightarrow$ independent choice logic (AIJ 1997).

Independent Choice Logic

- An alternative is a set of atomic formula.
  $\mathcal{C}$, the choice space is a set of disjoint alternatives.
- $\mathcal{F}$, the facts is a logic program that gives consequences of choices.
- P_0 a probability distribution over alternatives:

$$\forall A \in \mathcal{C} \sum_{a \in A} P_0(a) = 1.$$

Meaningless Example

$$\mathcal{C} = \{\{c_1, c_2, c_3\}, \{b_1, b_2\}\}$$

$$\begin{aligned}\mathcal{F} = \{\ & f \leftarrow c_1 \wedge b_1, && f \leftarrow c_3 \wedge b_2, \\ & d \leftarrow c_1, && d \leftarrow \neg c_2 \wedge b_1, \\ & e \leftarrow f, && e \leftarrow \neg d\}\end{aligned}$$

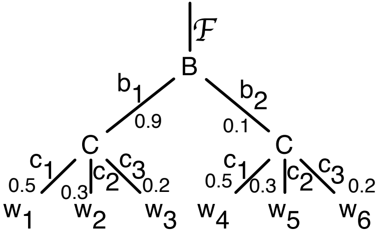
$$\begin{array}{lll} P_0(c_1) = 0.5 & P_0(c_2) = 0.3 & P_0(c_3) = 0.2 \\ P_0(b_1) = 0.9 & P_0(b_2) = 0.1 & \end{array}$$

Semantics of ICL

Probabilities are defined by a (possible infinite) semantic tree:

- Root has one choice corresponding to $\mathcal{F}$
- Each internal node corresponds to an alternative: child for each element of the alternative.

Meaningless Example: Semantics

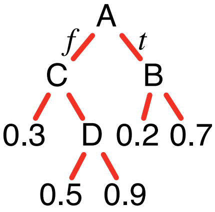


w_1	$\models$	c_1	b_1	f	d	e	$P(w_1) = 0.45$
w_2	$\models$	c_2	b_1	$\neg f$	$\neg d$	e	$P(w_2) = 0.27$
w_3	$\models$	c_3	b_1	$\neg f$	d	$\neg e$	$P(w_3) = 0.18$
w_4	$\models$	c_1	b_2	$\neg f$	d	$\neg e$	$P(w_4) = 0.05$
w_5	$\models$	c_2	b_2	$\neg f$	$\neg d$	e	$P(w_5) = 0.03$
w_6	$\models$	c_3	b_2	f	$\neg d$	e	$P(w_6) = 0.02$

$$P(e) = 0.45 + 0.27 + 0.03 + 0.02 = 0.77$$

Belief Networks, Decision trees and ICL rules

- There is a local mapping from belief networks into ICL.
- Rules can represent decision tree representation of conditional probabilities:

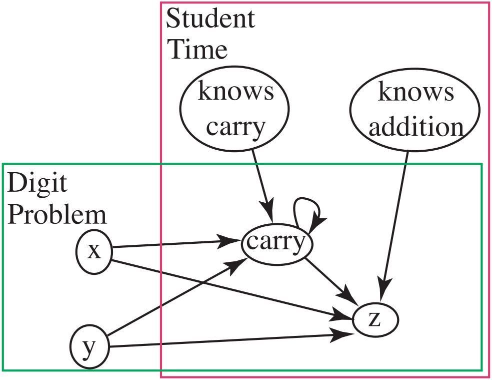


P(e|A,B,C,D)

$$
\begin{array}{ll}
e \leftarrow a \wedge b \wedge h_1. & P_0(h_1) = 0.7 \\
e \leftarrow a \wedge \neg b \wedge h_2. & P_0(h_2) = 0.2 \\
e \leftarrow \neg a \wedge c \wedge d \wedge h_3. & P_0(h_3) = 0.9 \\
e \leftarrow \neg a \wedge c \wedge \neg d \wedge h_4. & P_0(h_4) = 0.5 \\
e \leftarrow \neg a \wedge \neg c \wedge h_5. & P_0(h_5) = 0.3
\end{array}
$$

Example: Multi-digit addition

$$
\begin{array}{rcccc}
 & x_{j_x} & \cdots & x_2 & x_1 \\
+ & y_{j_z} & \cdots & y_2 & y_1 \\
\hline
 & z_{j_z} & \cdots & z_2 & z_1
\end{array}
$$



ICL rules for multi-digit addition

$$
\begin{aligned}
z(D,P,S,T) = V \leftarrow\; & \\
& x(D,P) = Vx \wedge \\
& y(D,P) = Vy \wedge \\
& carry(D,P,S,T) = Vc \wedge \\
& knowsAddition(S,T) \wedge \\
& \neg mistake(D,P,S,T) \wedge \\
& V \text{ is } (Vx + Vy + Vc) \text{ div } 10.
\end{aligned}
$$

$$
\begin{aligned}
z(D,P,S,T) = V \leftarrow\; & \\
& knowsAddition(S,T) \wedge \\
& mistake(D,P,S,T) \wedge \\
& selectDig(D,P,S,T) = V. \\
z(D,P,S,T) = V \leftarrow\; & \\
& \neg knowsAddition(S,T) \wedge \\
& selectDig(D,P,S,T) = V.
\end{aligned}
$$

Alternatives:

$\forall DPST \{noMistake(D,P,S,T), mistake(D,P,S,T)\}$

$\forall DPST \{selectDig(D,P,S,T) = V \mid V \in \{0..9\}\}$

Outline

1. Semantic Science Overview
   - Ontologies
   - Data
   - Theories
2. Representing Probabilistic Theories
   - Feature-based Theories
   - First-order probabilistic models
   - Probabilities with Ontologies
   - Existence and Identity Uncertainty
3. Pragmatics of Real Theories

Random Variables and Triples

- Reconcile:
  - random variables of probability theory
  - individuals, classes, properties of modern ontologies

Random Variables and Triples

- Reconcile:
  - random variables of probability theory
  - individuals, classes, properties of modern ontologies
- For functional properties:
  random variable for each $\langle \textit{individual}, \textit{property} \rangle$ pair, where the domain of the random variable is the range of the property.
- For non-functional properties:
  Boolean random variable for each $\langle \textit{individual}, \textit{property}, \textit{value} \rangle$ triple.

Triples and Probabilities

- $\langle individual, property, value\rangle$ triples are complete for representing relations
- $\langle individual, property, value, probability\rangle$ quadruples can represent probabilities of relations (or reify again)
- e.g., in addition $P(z(3, prob23, fred, t3) = 4) = 0.43$:

$$\left.\begin{array}{l}\langle z543, type, AdditionZValue\rangle \\ \langle z543, digit, 3\rangle \\ \langle z543, problem, prob23\rangle \\ \langle z543, student, fred\rangle \\ \langle z543, time, t3\rangle\end{array}\right\}\text{defines random variable}$$

$$\left.\begin{array}{l}\langle z543, valueWithProb, 4, 0.43\rangle \\ \langle z543, valueWithProb, 5, 0.03\rangle \\ \ldots\end{array}\right\}\text{defines distribution}$$

Probabilities and Aristotelian Definitions

Aristotelian definition

$$\begin{aligned} \mathit{ApartmentBuilding} \equiv\ & \mathit{ResidentialBuilding}\,\& \\ & \mathit{NumUnits} = \mathit{many}\,\& \\ & \mathit{Ownership} = \mathit{rental} \end{aligned}$$

leads to probability over property values

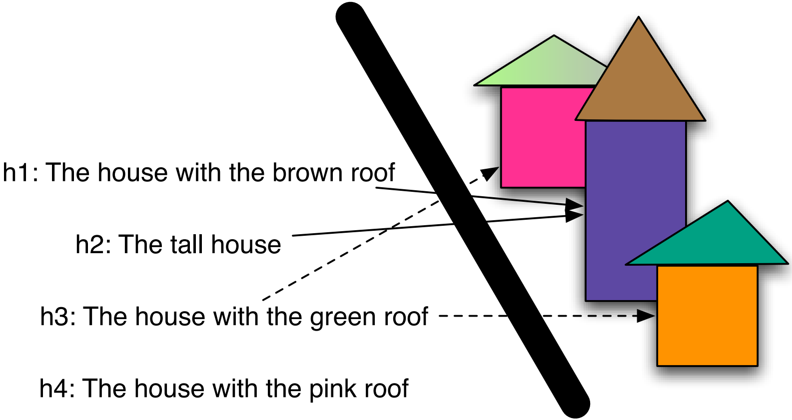
$$\begin{aligned} & P(\langle A, \mathit{type}, \mathit{ApartmentBuilding}\rangle) \\ & = P(\langle A, \mathit{type}, \mathit{ResidentialBuilding}\rangle) \times \\ & \quad P(\langle A, \mathit{NumUnits}, \mathit{many}\rangle \mid \langle A, \mathit{type}, \mathit{ResidentialBuilding}\rangle) \times \\ & \quad P(\langle A, \mathit{Ownership}, \mathit{rental}\rangle \mid \langle A, \mathit{NumUnits}, \mathit{many}\rangle, \\ & \qquad \langle A, \mathit{type}, \mathit{ResidentialBuilding}\rangle) \end{aligned}$$

No need to consider undefined propositions.

Outline

1. Semantic Science Overview
   - Ontologies
   - Data
   - Theories
2. Representing Probabilistic Theories
   - Feature-based Theories
   - First-order probabilistic models
   - Probabilities with Ontologies
   - Existence and Identity Uncertainty
3. Pragmatics of Real Theories

Existence and Identity



Clarity Principle

Clarity principle: probabilities must be over well-defined propositions.

- What if an individual doesn't exist?
  - $house(h4) \wedge roof\_colour(h4, pink) \wedge \neg exists(h4)$

Clarity Principle

Clarity principle: probabilities must be over well-defined propositions.

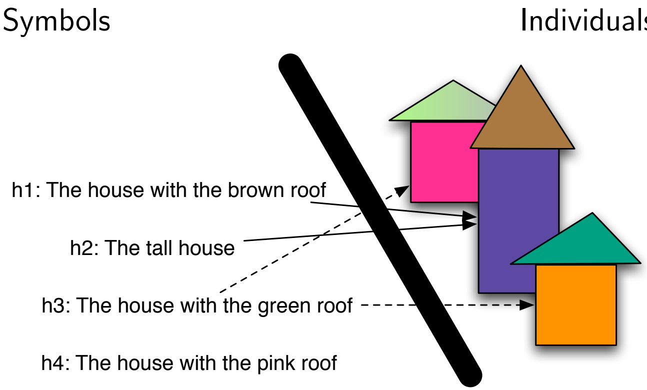
- What if an individual doesn't exist?
  - $house(h4) \wedge roof\_colour(h4, pink) \wedge \neg exists(h4)$
- What if more than one individual exists? Which one are we referring to?
  —In a house with three bedrooms, which is the second bedroom?

Clarity Principle

Clarity principle: probabilities must be over well-defined propositions.

- What if an individual doesn't exist?
  - $house(h4) \wedge roof\_colour(h4, pink) \wedge \neg exists(h4)$
- What if more than one individual exists? Which one are we referring to?
  —In a house with three bedrooms, which is the second bedroom?
- Reified individuals are special:
  — Non-existence means the relation is false.
  — Well defined what doesn't exist when existence is false.
  — Reified individuals with the same description are the same individual.

Correspondence Problem



c symbols and i individuals $\longrightarrow c^{i+1}$ correspondences

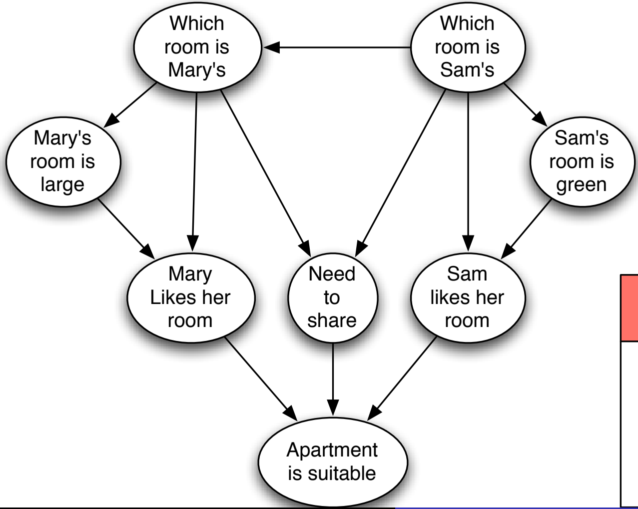
Role assignments

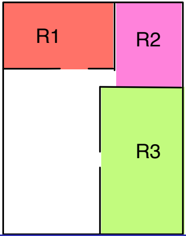
Theory about what apartment Mary would like.

Whether Mary likes an apartment depends on:

- Whether there is a bedroom for daughter Sam
- Whether Sam's room is green
- Whether there is a bedroom for Mary
- Whether Mary's room is large
- Whether they share

Role assignments

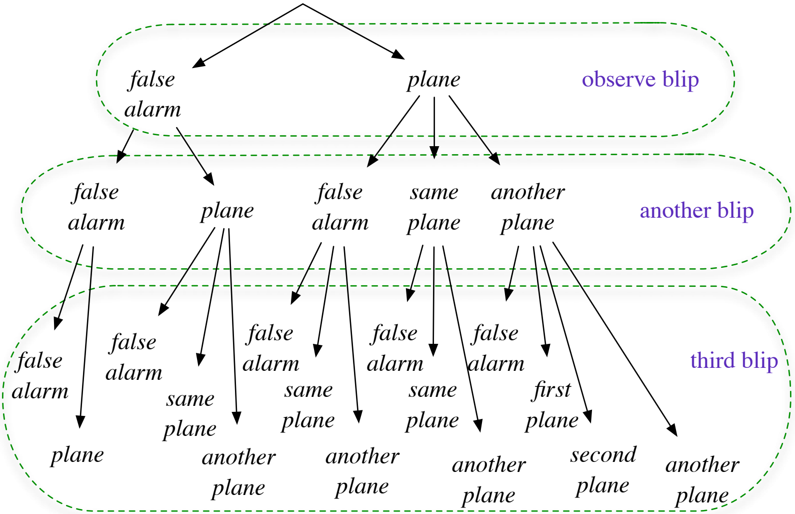




Number and Existence Uncertainty

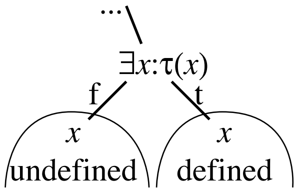
- PRMs (Pfeffer et al.), BLOG (Milch et al.): distribution over the number of individuals. For each number, reason about the correspondence.
- NP-BLOG (Carbonetto et al.): keep asking: is there one more?
  e.g., if you observe a radar blip, there are three hypotheses:
  - the blip was produced by plane you already hypothesized
  - the blip was produced by another plane
  - the blip wasn't produced by a plane

Existence Example



First-order Semantic Trees

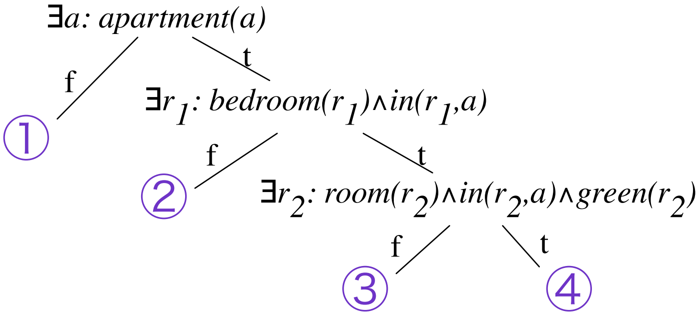
You can split on quantified first-order formulae:



- The "true" sub-tree is in the scope of x
- The "false" sub-tree is not in the scope of x

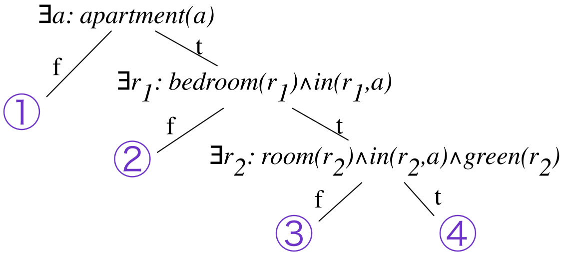
A logical generative model generates a first-order semantic tree.

First-order Semantic Tree (cont)



1. there is no apartment
2. there is no bedroom in the apartment
3. there is a bedroom but no green room
4. there is a bedroom and a green room

First-order Semantic Tree (cont)



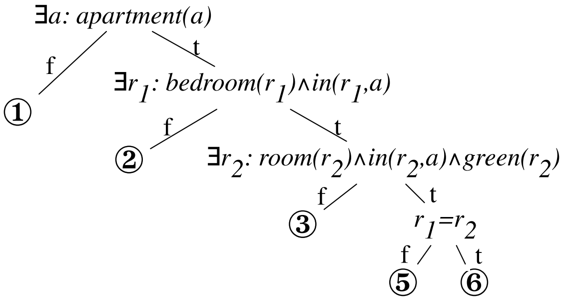
Path formulae:

① $(\neg \exists a \; apt(a))$

② $\exists a \; apt(a) \wedge \neg(\exists a \; apt(a) \wedge \exists r_1 \; br(r_1) \wedge in(r_1, a))$

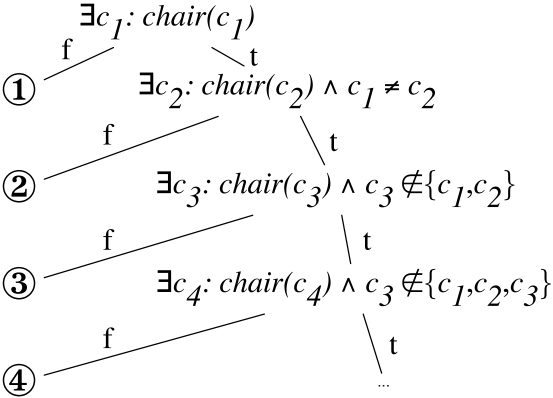
④ $\exists a \; apt(a) \wedge \exists r_1 \; br(r_1) \wedge in(r_1, a) \wedge \exists r_2 \; room(r_2) \wedge in(r_2, a) \wedge green(r_2)$

First-order Semantic Tree (cont)

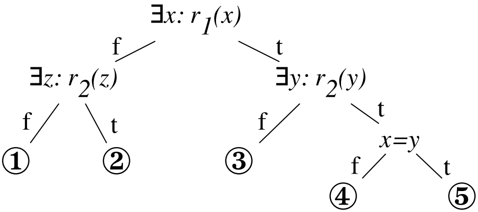


- ⑥ $\exists a\ apt(a) \wedge \exists r_1\ br(r_1) \wedge in(r_1, a) \wedge \exists r_2\ room(r_2) \wedge in(r_2, a) \wedge green(r_2) \wedge r_1 = r_2$
  There is a green bedroom.
- ⑤ There is a bedroom and a green room, but no green bedroom.

Distributions over number

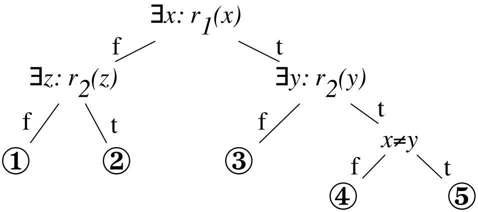


Roles and Identity (1)



1. there no individual filling either role
2. there is an individual filling role r_2 but none filling r_1
3. there is an individual filling role r_1 but none filling r_2
4. only different individuals fill roles r_1 and r_2
5. some individual fills both roles r_1 and r_2

Roles and Identity (2)



① there no individual filling either role

② there is an individual filling role r_2 but none filling r_1

③ there is an individual filling role r_1 but none filling r_2

④ only the same individual fill roles r_1 and r_2

⑤ there are different individuals that fill roles r_1 and r_2

Exchangeability

- First-order semantic trees can represent existence uncertainty, but not how to draw balls out of urns!

Exchangeability

- First-order semantic trees can represent existence uncertainty, but not how to draw balls out of urns!
- Consider definition of conditional probability:

$$P(h|e) = \frac{P(h \wedge e)}{P(e)}$$

What if h refers to an individual in e?

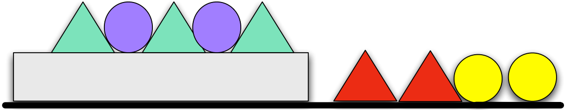
Exchangeability

- First-order semantic trees can represent existence uncertainty, but not how to draw balls out of urns!
- Consider definition of conditional probability:

$$P(h|e) = \frac{P(h \wedge e)}{P(e)}$$

  What if h refers to an individual in e?
- Exchangeability: a priori each individual is equally likely to be chosen.

Exchangeability

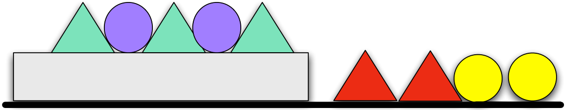


Consider the query:

$$P(green(x) \mid \exists x\ triangle(x) \wedge \exists y\ circle(y) \wedge touching(x, y))$$

The answer depends on how the x and y were chosen!

Protocol for Observing



$$P(green(x) \mid \exists x\ triangle(x) \wedge \exists y\ circle(y) \wedge touching(x, y))$$

$commit(x)$	$commit(y)$	$commit(x, y)$
$commit(y)$	$commit(x)$	
$3/4$	$2/3$	$4/5$

Outline

1. Semantic Science Overview
   - Ontologies
   - Data
   - Theories
2. Representing Probabilistic Theories
   - Feature-based Theories
   - First-order probabilistic models
   - Probabilities with Ontologies
   - Existence and Identity Uncertainty
3. Pragmatics of Real Theories

Expert Models

What if the models are provided by the experts in the field?

- not covering — only provide positive models
- not exclusive — they are often refinements of each other
- described at various levels of abstraction and detail
- often the experts don't know the probabilities and there is little data to estimate them

Providing Probabilities

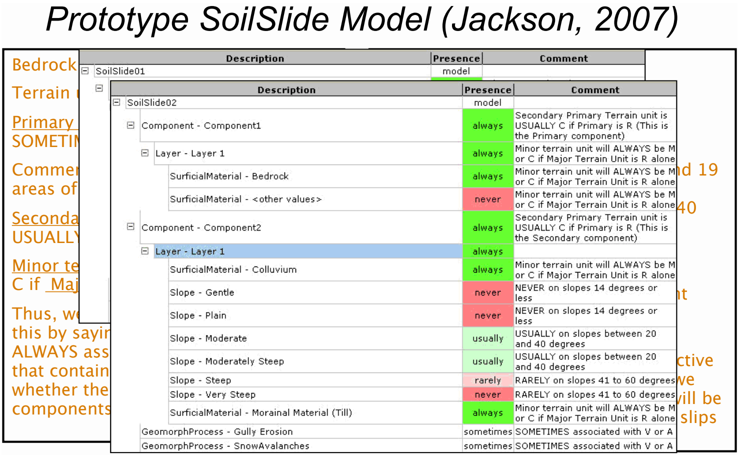
Experts are reluctant to give probabilities:

- No data from which to estimate them
- People who want to make decision use more information than provided in our theories
- Difficult to combine marginal probabilities with new information to make decisions
- It is *not* because decision theory is inappropriate. Decision makers use probabilities and utilities.

What we do

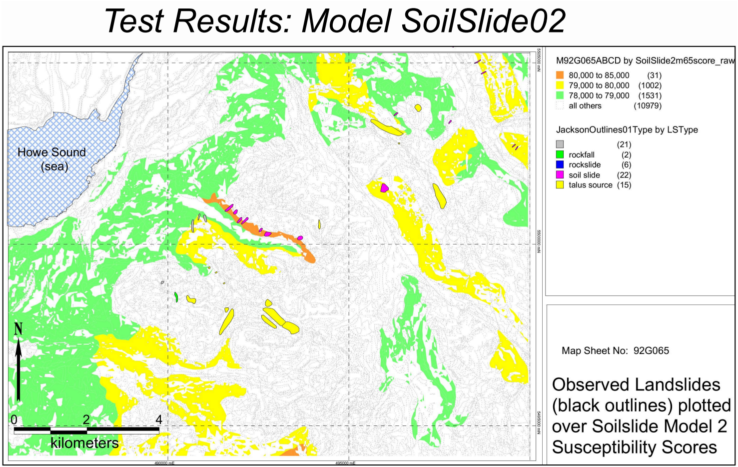
- Use qualitative probabilities: {*always*, *usually*, *sometimes*, *rarely*, *never*}.
- With thousands of instances and hundreds of models, find the most likely and the rationale.
- Independence assumptions.

Example Model



Description	Presence	Comment
SoilSlide02	model	
Component - Component1	always	Secondary Primary Terrain unit is USUALLY C if Primary is R (This is the Primary component)
Layer - Layer 1	always	Minor terrain unit will ALWAYS be M or C if Major Terrain Unit is R alone
SurficialMaterial - Bedrock	always	Minor terrain unit will ALWAYS be M or C if Major Terrain Unit is R alone
SurficialMaterial - <other values>	never	Minor terrain unit will ALWAYS be M or C if Major Terrain Unit is R alone
Component - Component2	always	Secondary Primary Terrain unit is USUALLY C if Primary is R (This is the Secondary component)
Layer - Layer 1	always	
SurficialMaterial - Colluvium	always	Minor terrain unit will ALWAYS be M or C if Major Terrain Unit is R alone
Slope - Gentle	never	NEVER on slopes 14 degrees or less
Slope - Plain	never	NEVER on slopes 14 degrees or less
Slope - Moderate	usually	USUALLY on slopes between 20 and 40 degrees
Slope - Moderately Steep	usually	USUALLY on slopes between 20 and 40 degrees
Slope - Steep	rarely	RARELY on slopes 41 to 60 degrees
Slope - Very Steep	never	RARELY on slopes 41 to 60 degrees
SurficialMaterial - Morainal Material (Till)	always	Minor terrain unit will ALWAYS be M or C if Major Terrain Unit is R alone
GeomorphProcess - Gully Erosion	sometimes	SOMETIMES associated with V or A
GeomorphProcess - SnowAvalanches	sometimes	SOMETIMES associated with V or A

Example Model



Conclusion

- Semantic science is a way to develop and deploy knowledge about how the world works.
- Scientists (and others) develop theories that refer to standardized ontologies and predict for new cases.
- Multiple theories—forming theory ensembles—are needed to make predictions in particular cases.
- For each prediction, we want to be able to ask, what theories it is based on.
- For each theory, we want to be able to ask what evidence it is based on.
- This talk is deliberately pre-theoretic. Many formalisms will be developed and discarded before we converge on useful representations.

To Do

- Theories of combining theories.
- Representing, reasoning and learning complex (probabilistic) theories.
- Build infrastructure to allow publishing and interaction of ontologies, data, theories, theory ensembles, evaluation criteria, meta-data.
- Build inverse semantic science web:
  - Given a theory, find relevant data
  - Given data, find theory ensembles that make predictions on the data
  - Given a new case, find relevant theory ensembles with explanations
- More complex models, e.g., for relational reinforcement learning where individuals are created and destroyed