# Probabilistic Relational Learning and Inductive Logic Programming at a Global Scale

David Poole

Department of Computer Science,
University of British Columbia

29 June 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]

# Questions considered

- [Grad student:] All of the easy problems have been solved; how do I choose a what to work on?

# Questions considered

- [Grad student:] All of the easy problems have been solved; how do I choose a what to work on?
- Search engines give results of my query, but why should I believe the answers?

# Questions considered

- [Grad student:] All of the easy problems have been solved; how do I choose a what to work on?
- Search engines give results of my query, but why should I believe the answers?
- The semantic web is supposed to make human knowledge accessible to computers, but how can we evaluate that knowledge and go beyond the sum of human knowledge?

# Questions considered

- [Grad student:] All of the easy problems have been solved; how do I choose a what to work on?
- Search engines give results of my query, but why should I believe the answers?
- The semantic web is supposed to make human knowledge accessible to computers, but how can we evaluate that knowledge and go beyond the sum of human knowledge?
- How is (probabilistic) inductive logic programming central to the semantic web?

# Questions considered

- [Grad student:] All of the easy problems have been solved; how do I choose a what to work on?
- Search engines give results of my query, but why should I believe the answers?
- The semantic web is supposed to make human knowledge accessible to computers, but how can we evaluate that knowledge and go beyond the sum of human knowledge?
- How is (probabilistic) inductive logic programming central to the semantic web?
- What will AI and the web look like in 2025?

# History of AI — a perspective from 2025

- Semantic web has evolved into the world-wide mind (WWM) — a distributed repository of all knowledge, backed up by the best science available.
- The world-wide mind doesn't just accept new knowledge but critically evaluates it and generates new knowledge.
- Scientists freed from mundane data analysis, develop new hypotheses, interesting questions, and observational data.
- World-wide mind is the expert on all questions of truth and makes the best predictions. (Using hypotheses provided by a mix of humans and machine learning).
- Public discourse on values (utilities) to determine the best course of actions for individuals, organizations and society.

# Finding information; e.g. diagnosis from symptoms

| 2010 | 2025 |
|---|---|
| • need to guess keywords; re-guess until exhaustion | • keywords + context + ontologies → unambiguous query |

# Finding information; e.g. diagnosis from symptoms

| 2010 | 2025 |
| --- | --- |
| • need to guess keywords; re-guess until exhaustion | • keywords + context + ontologies $\rightarrow$ unambiguous query |
| • what information found is based on popularity and/or appeal to authority | |

# Finding information; e.g. diagnosis from symptoms

| 2010 | 2025 |
|---|---|
| • need to guess keywords; re-guess until exhaustion | • keywords + context + ontologies $\rightarrow$ unambiguous query |
| • what information found is based on popularity and/or appeal to authority | • information based on best evidence available in world |

# Finding information; e.g. diagnosis from symptoms

| 2010 | 2025 |
| --- | --- |
| • need to guess keywords; re-guess until exhaustion | • keywords + context + ontologies → unambiguous query |
| • what information found is based on popularity and/or appeal to authority | • information based on best evidence available in world |
| • verify information based on other sites (with different wording) | • information justified by presenting the evidence for and against it |

# Finding information; e.g. diagnosis from symptoms

| 2010 | 2025 |
|---|---|
| • need to guess keywords; re-guess until exhaustion | • keywords + context + ontologies → unambiguous query |
| • what information found is based on popularity and/or appeal to authority | • information based on best evidence available in world |
| • verify information based on other sites (with different wording) | • information justified by presenting the evidence for and against it |
| • extract information from text and graphics to make decisions | • decisions based on evidence and utilities |

# Believing information

| 2010 | 2025 |
|---|---|
| • skeptics throw doubt on science and scientists say "trust us" | • data is available for all to view; all alternative hypotheses can be evaluated |

# Believing information

| 2010 | 2025 |
|---|---|
| • skeptics throw doubt on science and scientists say "trust us" | • data is available for all to view; all alternative hypotheses can be evaluated |
| • politicians campaign on what is true and what they will do | • politicians campaign on their values |

# Believing information

| 2010 | 2025 |
|---|---|
| • skeptics throw doubt on science and scientists say "trust us" | • data is available for all to view; all alternative hypotheses can be evaluated |
| • politicians campaign on what is true and what they will do | • politicians campaign on their values |
| • food shopping is based on price and brands | • food shopping based on optimizing health and well-being (users goals and values, and known risks) |

# AI Research

| 2010 | 2025 |
|---|---|
| • separation of uncertainty and KR issues<br>— ML ignores ontologies<br>— rich representations ignore uncertainty | • uncertainty and ontologies are integral parts of world-wide mind |
| • semantic web in its infancy | • world wide mind being used |
| • relational representations starting to be used in ML | • rich representations with uncertainty ubiquitous |
| • learning based on one or few homogeneous data sets | • learning from all data in world |
| • data sets usable only by specialists | • data sets published, available, persistent and interoperable |

# Outline

1. Semantic Science Overview
   - Ontologies
   - Data
   - Hypotheses and Theories
   - Models
2. Feature-based model construction
3. Lifted Inference in Relational Domains
4. Existence and Identity Uncertainty

# Notational Minefield

- Variable (probability, logic, programming languages)
- Model (science, probability, logic, fashion)
- Parameter (mathematics, statistics)
- Domain (science, logic, probability, mathematics)
- Object/class (object-oriented programming, ontologies)
- $=$ (probability, logic)
- First-order (logic, dynamical systems)

# Science is the foundation of belief

- If a KR system makes a prediction, we should ask: what evidence is there? The system should be able to provide such evidence.
- A knowledge-based system should believe based on evidence. Not all beliefs are equally valid.
- The mechanism that has been developed for judging knowledge is called science. We trust scientific conclusions because they are based on evidence.

# Science is the foundation of belief

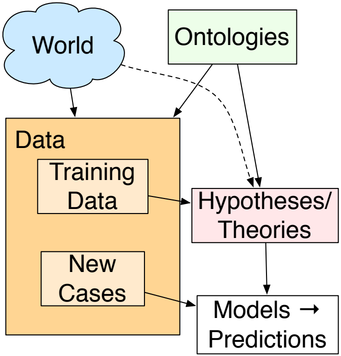
- If a KR system makes a prediction, we should ask: what evidence is there? The system should be able to provide such evidence.
- A knowledge-based system should believe based on evidence. Not all beliefs are equally valid.
- The mechanism that has been developed for judging knowledge is called science. We trust scientific conclusions because they are based on evidence.
- The semantic web is an endeavor to make all of the world's knowledge accessible to computers.
- We have used to term semantic science, in an anaolgous way to the *semantic web*.
- Claim: semantic science will form the foundation of the world-wide mind.

# Science as the foundation of world-wide mind

*Science* can be about anything:

- 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
- which celebrities are having affairs

# Semantic Science



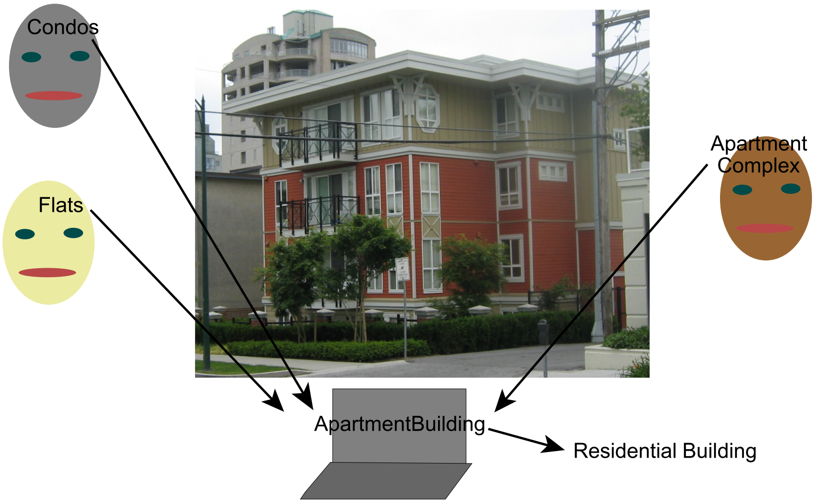

- Ontologies represent the meaning of symbols.
- Data that adheres to ontologies are published.
- Hypotheses that make (probabilistic) predictions on data are published.
- Data used to evaluate hypotheses; the best hypotheses are theories.
- Hypotheses form models for predictions on new cases.
- All evolve in time.

# Outline

# 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

# 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

$\langle Individual, Property, Value \rangle$ triples are universal representations of relations.

# Semantic Web Ontology Languages

- URI — universal resource identifier; everything is a resource
- RDF — language for triples in XML
- RDF Schema — define resources in terms of each other: class, type, subClassOf, subPropertyOf, collections...
- OWL — defines vocabulary for individuals, properties and classes: equality, restricting domains and ranges of properties, transitivity, cardinality...

# 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**.

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

  *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 are defined in terms of properties.

# Outline

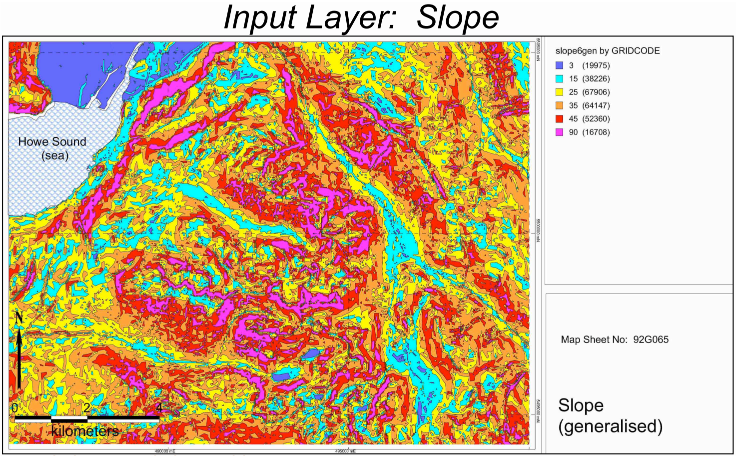
1. Semantic Science Overview
   - Ontologies
   - Data
   - Hypotheses and Theories
   - Models
2. Feature-based model construction
3. Lifted Inference in Relational Domains
4. Existence and Identity Uncertainty

# Data

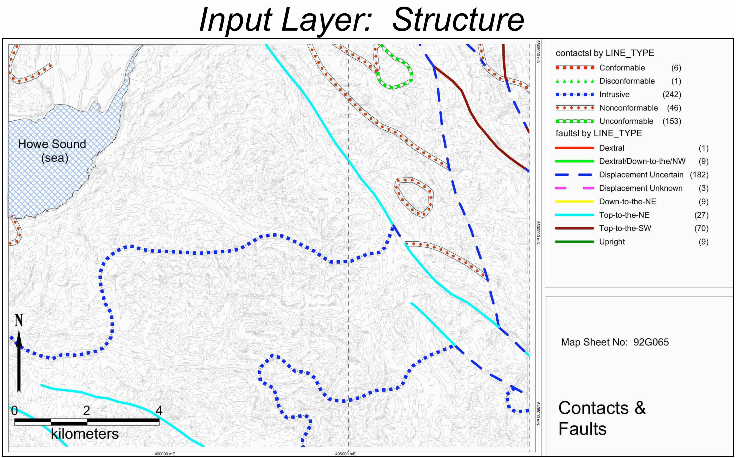
Real data is messy!

- 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/

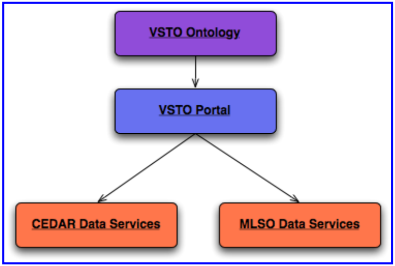
**Welcome to the Virtual Solar Terrestrial Observatory**

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 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

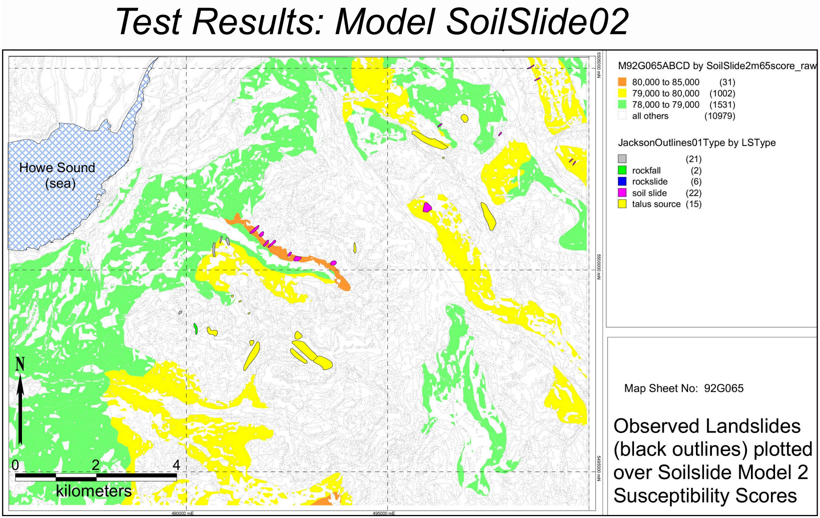
1. Semantic Science Overview
   - Ontologies
   - Data
   - Hypotheses and Theories
   - Models
2. Feature-based model construction
3. Lifted Inference in Relational Domains
4. Existence and Identity Uncertainty

# Hypotheses make predictions on data

Hypotheses are procedures that make prediction on data.
Theories are hypotheses that best fit the observational data.

- Hypotheses 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 hypotheses (e.g., taking into account simplicity and elegance)
- Semantic science search engine: extract theories from published hypotheses.

# Example Prediction from a Hypothesis



# Applying hypotheses to new cases

- Hypotheses are often narrow, e.g., prognosis of people with a lung cancer.
- Hypotheses are general in the sense that they can be adapted to different cases.
- A model is a set of hypotheses applied to a particular case.
  - Judge hypotheses by how well they fit into models.
  - Models can be judged by simplicity.
  - Hypothesis designers don't need to game the system by manipulating the generality of hypotheses

# Dynamics of Semantic Science

- New data and hypotheses are continually added.
- 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.
- Hypotheses engineered + learned (e.g., using ILP)
- Ontologies evolve with hypotheses:
  A hypothesis learns useful unobserved features
  $\longrightarrow$ 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 hypotheses that use them
  — role of a vocabulary is to describe useful distinctions.

# Levels of Semantic Science

0. Deterministic semantic science where all of the hypotheses 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

# Feature-Based Semantic Science

- World is described in terms of features and values.
- Random variables / features correspond to properties.
- Random variables / features are not defined in all contexts.
- Aristotelian definitions: each class is defined in terms of
    - genus (superclass) and
    - differentia (property restrictions that distinguish this class).
- Conditioning on a class means observing its differentia are true

# Example Ontology

| Property | Domain | Range |
|---|---|---|
| Age | person | integer |
| Sex | person | {male,female} |
| HasLump | person | boolean |
| LumpShape | personWithLump | {circular, oblong, irregular} |
| LumpLocn | personWithLump | {leg,torso,arm,head} |
| CancLump | personWithLump | boolean |
| LumpColour | personWithLump | {red,pink,brown,...} |
| HasCancer | person | boolean |
| HasLungCancer | personWithCancer | boolean |
| OutcomeAtYear | person | {well,sick,dead} |

| Class | Genus | Differentia |
|---|---|---|
| person | thing | IsPerson=true |
| personWithLump | person | HasLump=true |
| personWithCancer | person | HasCancer=true |

# Partial Ontology in OWL

```
FunctionalDataProperty(HasLump)
DataPropertyDomain(HasLump person)
DataPropertyRange(HasLump xsd:boolean)
EquivalentClasses(personWithLump
           DataHasValue(HasLump true))
FunctionalDataProperty(CancLump)
DataPropertyDomain(CancLump personWithLump)
DataPropertyRange(CancLump xsd:boolean)
ObjectPropertyDomain(LumpShape personWithLump)
ObjectPropertyRange(LumpShape
   ObjectOneOf(circular oblong irregular))
SubClassOf(DataHasValue(CancLump true)
    personWithCancer)
```

# Observational Data

A data set refers to a set of imported ontologies and consists of a set of $\langle c, O, t \rangle$ triples, where:

- the context $c$ in which the data was collected
- the features $O$ that were observed
- a table $t$ on $O$

The symbols used in the data set are defined in the ontologies.
[This talk will ignore interventions.]

# Example Data

person:

| Age | Sex | Coughs | HasLump |
|---|---|---|---|
| 23 | male | true | true |
| ... | ... | ... | ... |

personWithLump:

| LumpLocn | LumpShape | LumpColour | CancLump |
|---|---|---|---|
| leg | oblong | red | false |
| ... | ... | ... | ... |

personWithCancer:

| HasLungCancer | TakenHerb | Age | OutcomeAtYear |
|---|---|---|---|
| true | true | 77 | dead |
| ... | ... | ... | ... |

# Hypotheses

A hypothesis makes predictions about some feature values.

A hypothesis $h$ is $\langle c, I, O, P \rangle$ where:

- $c$, a context, is a proposition that specifies when $h$ can be applied.
- $I$ is a set of input features about which $h$ does not make predictions
- $O$ is a set output features to predict (as a function of the input features).
- $P$ is a program to compute the output from the input (e.g., in ProbLog or Figaro)

Represents:

$$P(O|c, I)$$

# Example Hypotheses

- $H_1$ predicts the prognosis of people with lung cancer as a function of treatment:

$$\langle personWithCancer \wedge HasLungCancer, \{Treatment\}, \{OutcomeAtYear\}, P_1\rangle$$

- $H_2$ predicts the prognosis of people with cancer.

$$\langle personWithCancer, \{\}, \{OutcomeAtYear\}, P_2\rangle$$

- $H_3$ is a null hypothesis for the prognosis of people:

$$\langle person, \{\}, \{OutcomeAtYear\}, P_3\rangle$$

- $H_4$ predicts the prognosis of people with cancer, as a function of their income and age:

$$\langle personWithCancer, \{Income, Age\}, \{OutcomeAtYear\}, P_4\rangle$$

# Example Hypotheses (cont.)

- $H_5$ predicts whether people with cancer have lung cancer, as a function of coughing.

  $$\langle personWithCancer, \{Coughs\}, \{HasLungCancer\}, P_5\rangle$$

- $H_6$ predicts whether people have cancer.

  $$\langle person, \{\}, \{HasCancer\}, P_6\rangle$$

- $H_7$ and $H_8$ predict the shape of lumps as a function of whether they have cancer.

  $$\langle person, \{HasCancer\}, \{HasLump\}, P_7\rangle$$

  $$\langle personWithLump, \{HasCancer\}, \{LumpShape\}, P_8\rangle$$

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

# Models

To make a prediction given a query for a particular case, multiple hypotheses are used together in a model.

Each hypothesis in the model is used to predict a subset of its output features given its inputs.

A model $M$ needs to satisfy the following properties:

- $M$ is coherent: it does not rely on the value of a feature in a context where the feature is not defined
- $M$ is consistent: it does not make different predictions for any feature in any context.
- $M$ is predictive: it predicts the input of each hypothesis and predicts the query in every context that is possible.
- $M$ is minimal.

# Prototype Feature-based Model

A hypothesis instance is a tuple of the form $\langle t, c, I, O \rangle$ such that:

- $t$ is a hypothesis,
- $c$ a context in which the hypothesis is used
- $I$ a set of inputs used by the hypothesis
- $O$ a set of outputs the hypothesis is used to predict.

A model is a set of hypothesis instances that is coherent, consistent, predictive and minimal.

# Example

- $H_1$ predicts the prognosis of people with lung cancer as a function of treatment.
- $H_2$ predicts the prognosis of people with cancer.
- $H_3$, a null hypothesis, predicts the prognosis of people.
- $H_5$ predicts whether people with cancer have lung cancer, as a function of coughing.
- $H_6$ predicts whether people have cancer.

A possible model for
$P(OutcomeAtYear | person \wedge coughs \wedge treatment = none)$:
$\langle H_3, person \wedge \neg hasCancer, \{\}, \{OutcomeAtYear\}\rangle$
$\langle H_1, person \wedge hasLungCancer, \{Treatment\}, \{OutcomeAtYear\}\rangle$
$\langle H_2, person \wedge hasCancer \wedge \neg hlc, \{\}, \{OutcomeAtYear\}\rangle$
$\langle H_6, person, \{\}, \{HasCancer\}\rangle$
$\langle H_5, person \wedge hasCancer, \{Coughs\}, \{HasLungCancer\}\rangle$

# Feature-based predictions

- Multiple competing and complementary hypotheses
- Hypotheses can be judged and learned using all existing data sets
- A semantic science search engine would find best hypotheses (the theories) for any predictions.
- For a specific case, multiple hypotheses are combined to form a model.
- Each prediction relies on multiple hypotheses; each hypothesis is judged by multiple data sets.

# Outline

# Relational Learning

- Often the values of properties are not meaningful values but names of individuals.
- It is the properties of these individuals and their relationship to other individuals that needs to be learned.
- Relational learning has been studied under the umbrella of "Inductive Logic Programming" as the representations are often logic programs.

# Example: trading agent

What does Joe like?

| Individual | Property | Value |
|---|---|---|
| *joe* | *likes* | *resort_14* |
| *joe* | *dislikes* | *resort_35* |
| ... | ... | ... |
| *resort_14* | *type* | *resort* |
| *resort_14* | *near* | *beach_18* |
| *beach_18* | *type* | *beach* |
| *beach_18* | *covered_in* | *ws* |
| *ws* | *type* | *sand* |
| *ws* | *color* | *white* |
| ... | ... | ... |

Values of properties may be meaningless names.

# Example: trading agent

Possible theory that could be learned:

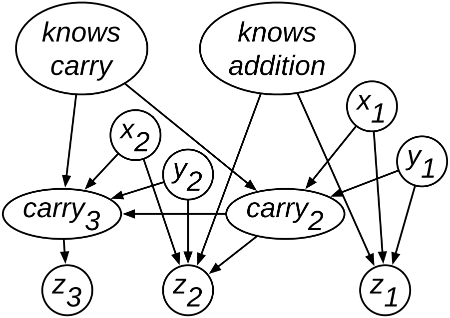
$$
\begin{aligned}
& prop(joe, likes, R) \leftarrow \\
& \qquad prop(R, type, resort) \wedge \\
& \qquad prop(R, near, B) \wedge \\
& \qquad prop(B, type, beach) \wedge \\
& \qquad prop(B, covered\_in, S) \wedge \\
& \qquad prop(S, type, sand).
\end{aligned}
$$

Joe likes resorts that are near sandy beaches.

- But we want probabilistic predictions.

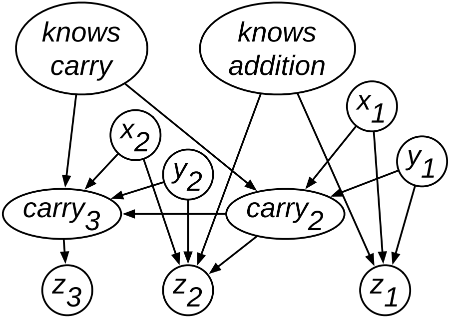
# Example Predicting students errors

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



# Example Predicting students errors

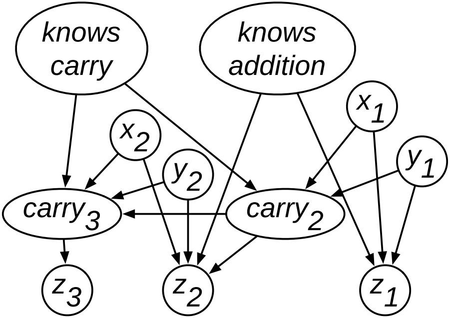
$$\begin{array}{cccc} & & x_2 & x_1 \\ + & & y_2 & y_1 \\ \hline & z_3 & z_2 & z_1 \end{array}$$




What if there were multiple digits

# Example Predicting students errors

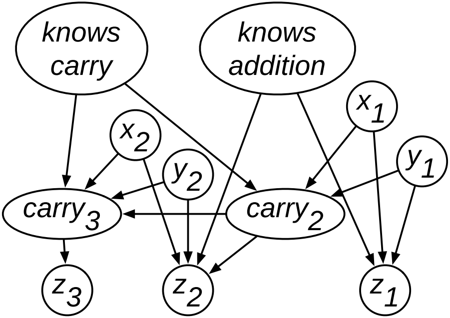
$$\begin{array}{rccc} & & x_2 & x_1 \\ + & & y_2 & y_1 \\ \hline & z_3 & z_2 & z_1 \end{array}$$




What if there were multiple digits, problems

# Example Predicting students errors

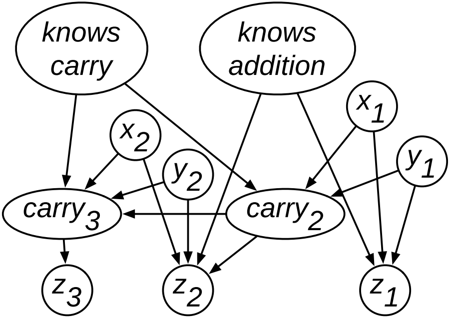
$$\begin{array}{rrrr} & & x_2 & x_1 \\ + & & y_2 & y_1 \\ \hline & z_3 & z_2 & z_1 \end{array}$$




What if there were multiple digits, problems, students

# Example Predicting students errors

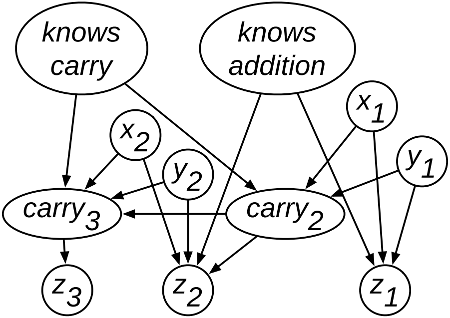
$$\begin{array}{rrrr} & & x_2 & x_1 \\ + & & y_2 & y_1 \\ \hline & z_3 & z_2 & z_1 \end{array}$$




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

# Example Predicting students errors

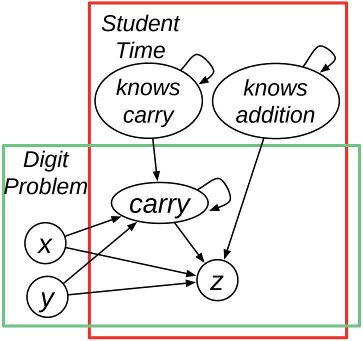
$$
\begin{array}{cccc}
 & & x_2 & x_1 \\
+ & & y_2 & y_1 \\
\hline
 & z_3 & z_2 & z_1
\end{array}
$$




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

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




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 $T$

☛ parametrized random variables

# Exchangeability

- Before we know anything about individuals, they are indistinguishable, and so should be treated identically.

# Independent Choice Logic (ICL)

- A language for relational probabilistic models.
- Idea: combine logic and probability, where all uncertainty in handled in terms of Bayesian decision theory, and logic specifies consequences of choices.
- An ICL theory consists of a choice space with probabilities over choices and a logic program that gives consequences of choices.

# ICL rules for multi-digit addition

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

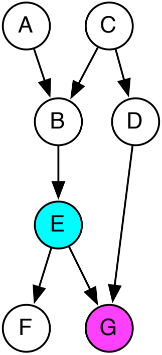
$$
\begin{aligned}
z(D,P,S,T) = V \leftarrow&\\
&knowsAddition(S,T) \wedge\\
&mistake(D,P,S,T) \wedge\\
&selectDig(D,P,S,T) = V.\\
z(D,P,S,T) = V \leftarrow&\\
&\neg knowsAddition(S,T) \wedge\\
&selectDig(D,P,S,T) = V.
\end{aligned}
$$

Alternatives:

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

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

# Bayesian Network Inference




$$P(E|g) = \frac{P(E \wedge g)}{p(g)}$$

$$P(E \wedge g) = \sum_F \sum_B \sum_C \sum_A \sum_D P(A)P(B|AC)$$
$$P(C)P(D|C)P(E|B)P(F|E)P(g|ED)$$
$$= \left(\sum_F P(F|E)\right)$$
$$\sum_B P(e|B) \sum_C P(C) \left(\sum_A P(A)P(B|AC)\right)$$
$$\left(\sum_D P(D|C)P(g|ED)\right)$$

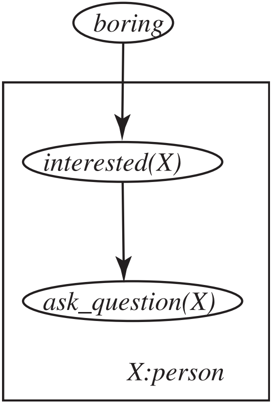
# Lifted Inference

- Idea: treat those individuals about which you have the same information as a block; just count them.
- Use the ideas from lifted theorem proving - no need to ground.
- Potential to be exponentially faster in the number of non-differentialed individuals.
- Relies on knowing the number of individuals (the population size).

# Parametrized belief networks

- Allow random variables to be parametrized. $interested(X)$
- Parameters correspond to logical variables. $X$
- Each parameter is typed with a population. $X : person$
- Each population has a size. $|person| = 1000000$
- Parametrized belief network means its grounding: for each combination of parameters, an instance of each random variable for each member of parameters' population. $interested(p_1) \ldots interested(p_{1000000})$
- Instances are independent (but can have common ancestors and descendants).

# Example parametrized belief network




$P(boring)$
$\forall X \; P(interested(X)|boring)$
$\forall X \; P(ask\_question(X)|interested(X))$

# First-order probabilistic inference

$$
\begin{array}{ccc}
\text{Parametrized Belief Network} & \xrightarrow{\text{FOVE}} & \text{Parametrized Posterior} \\
\Big\downarrow \text{ground} & & \Big\downarrow \text{ground} \\
\text{Belief Network} & \xrightarrow{\text{VE}} & \text{Posterior}
\end{array}
$$

# Theorem Proving and Unification

In 1965, Robinson showed how unification allows many ground steps with one step:

$$\underbrace{f(X,Z) \vee p(X,a) \qquad \neg p(b,Y) \vee g(Y,W)}_{f(b,Z) \vee g(a,W)}$$

Substitution $\{X/b, Y/a\}$ is the most general unifier of $p(X,a)$ and $p(b,Y)$.

# Variable Elimination and Unification

- Multiplying parametrized factors:

$$\underbrace{[f(X,Z),p(X,a)] \quad \times \quad [p(b,Y),g(Y,W)]}_{[f(b,Z),p(b,a),g(a,W)]}$$

  Doesn't work because the first parametrized factor can't subsequently be used for $X = b$ but can be used for other instances of $X$.

- We split $[f(X,Z),p(X,a)]$ into

$$[f(b,Z),p(b,a)]$$
$$[f(X,Z),p(X,a)] \text{ with constraint } X \neq b,$$

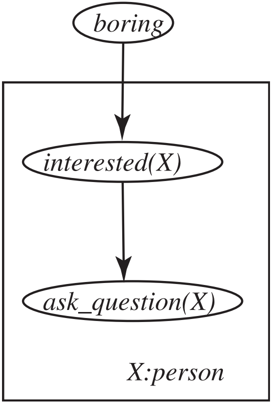
# Parametric Factors

A parametric factor is a triple $\langle C, V, t \rangle$ where

- $C$ is a set of inequality constraints on parameters,
- $V$ is a set of parametrized random variables
- $t$ is a table representing a factor from the random variables to the non-negative reals.

$$\left\langle \{X \neq sue\}, \{interested(X), boring\}, \begin{array}{|ll|l|} \hline interested & boring & Val \\ \hline yes & yes & 0.001 \\ yes & no & 0.01 \\ & \ldots & \\ \hline \end{array} \right\rangle$$

# Removing a parameter when summing




$n$ people

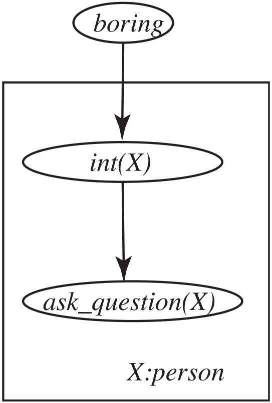
we observe no questions

Eliminate *interested*:

$$\langle \{\}, \{boring, interested(X)\}, t_1 \rangle$$
$$\langle \{\}, \{interested(X)\}, t_2 \rangle$$
$$\downarrow$$
$$\langle \{\}, \{boring\}, (t_1 \times t_2)^n \rangle$$

$(t_1 \times t_2)^n$ is computed pointwise; we can compute it in time $O(\log n)$.

# Counting Elimination




$|people| = n$

[de Salvo Braz et al. 2007] and [Milch et al. 08]

Eliminate *boring*:
VE: factor on $\{int(p_1), \ldots, int(p_n)\}$
Size is $O(d^n)$ where $d$ is size of range of interested.

Exchangeable: only the number of interested individuals matters.

Counting Formula:

| #interested | Value |
|---|---|
| 0 | $v_0$ |
| 1 | $v_1$ |
| ... | ... |
| n | $v_n$ |

Complexity: $O(n^{d-1})$.

# Potential of Lifted Inference

- Reduce complexity:

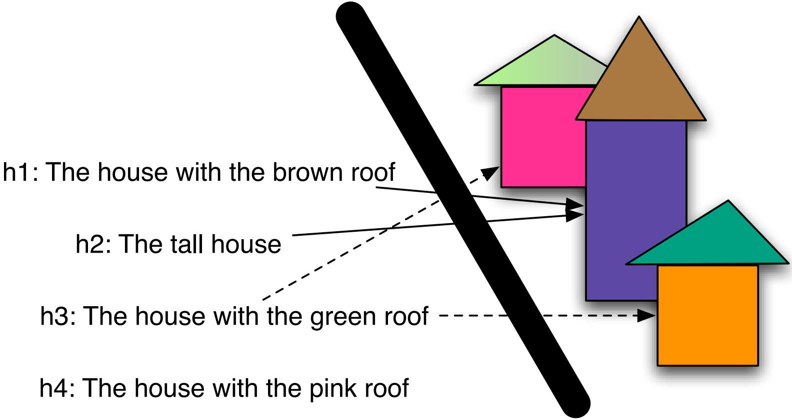
$$\textit{polynomial} \longrightarrow \textit{logarithmic}$$

$$\textit{exponential} \longrightarrow \textit{polynomial}$$

- We need a representation for the intermediate (lifted) factors that is closed under multiplication and summing out (lifted) variables.
- Still an open research problem.

# Outline

# 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.

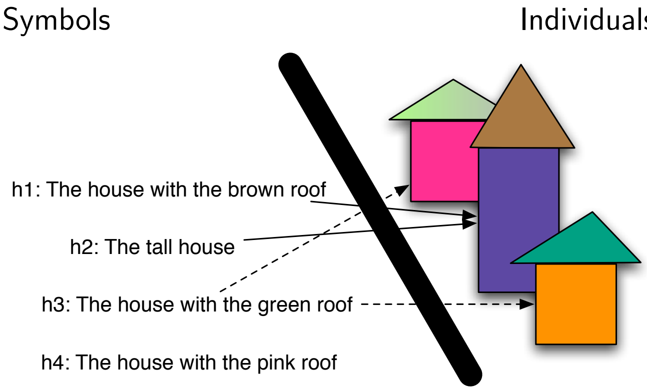
- What if an individual doesn't exist?
  - $house(h4) \wedge roof\_colour(h4, pink) \wedge \neg exists(h4)$
- What if more than one individual exists? Which one are we referring to?
  —In a house with three bedrooms, which is the second bedroom?

# 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

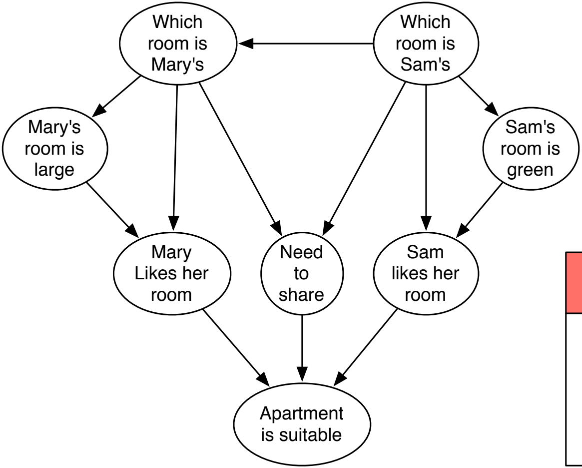
# Role assignments

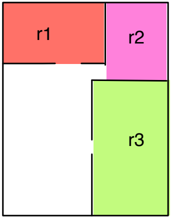
Hypothesis about what apartment Mary would like.

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

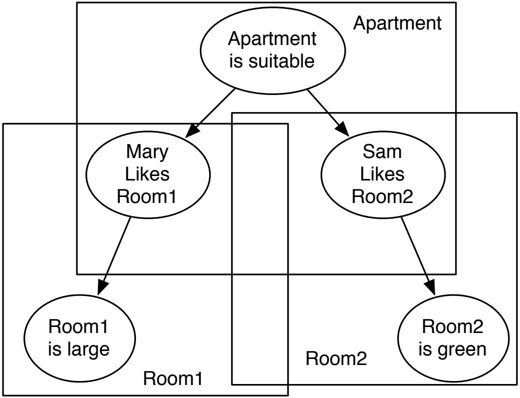
# BN Representation

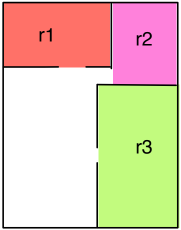



How can we condition on the observation of the apartment?

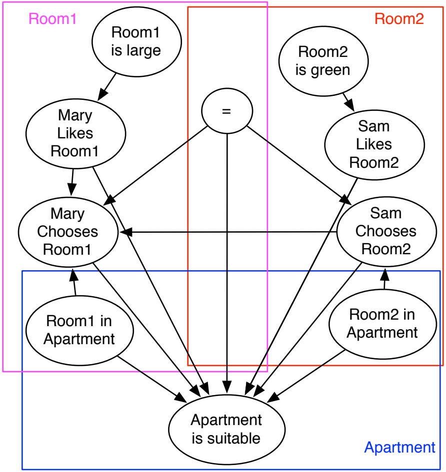
# Naive Bayes representation

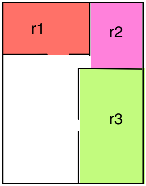



How do we specify that Mary chooses a room?
What about the case where they (have to) share?

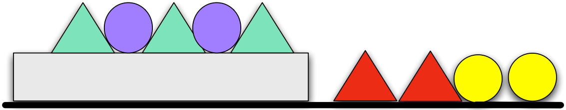
# Causal representation




How do we specify that Sam and Mary choose one room each, but they can like many rooms?

# Observation Protocols

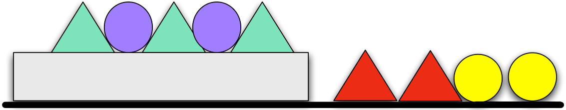


Observe a triangle and a rectangle touching. What is the probability the triangle is green?

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

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

# Protocol for Observing



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

| | | |
|---|---|---|
| $select(x)$ | $select(y)$ | $select(x, y)$ |
| $select(y)$ | $select(x)$ | |
| $3/4$ | $2/3$ | $4/5$ |

# Conclusion

- Semantic science is a way to develop and deploy knowledge about how the world works.
- Scientists (and others) develop hypotheses that refer to standardized ontologies and predict for new cases.
- For each prediction: what hypotheses it is based on?
- For each hypothesis: what evidence it is based on?
- Three subproblems:
    - Making predictions in a specific case.
    - Efficient inference in relational domains
    - Existence uncertainty and roles
- (Probabilistic) inductive logic programming is a core technology in constructing hypotheses.

# To Do

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