Logic, Probability and Computation: Statistical Relational AI and Beyond

David Poole

Department of Computer Science,
University of British Columbia

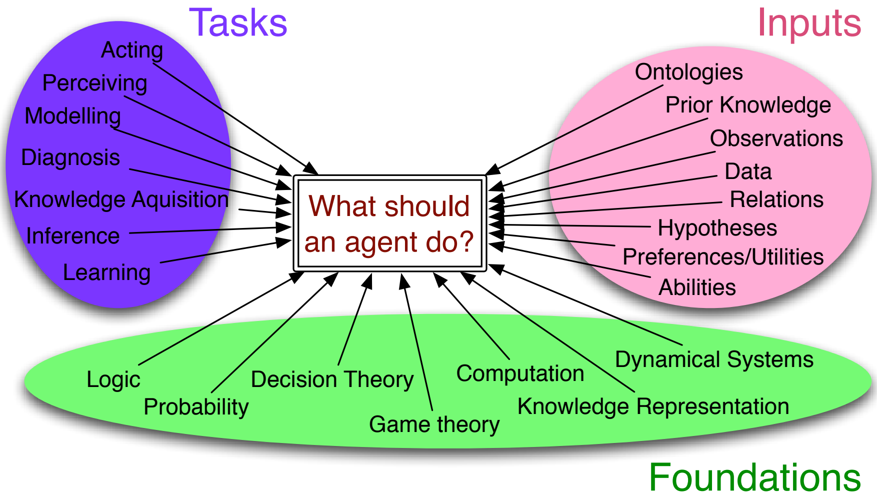
18 May 2011

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. . .

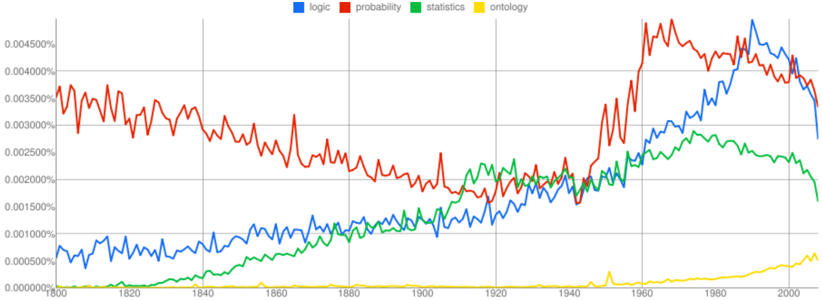
And if someone would doubt my results, I should say to him: "Let us calculate, Sir," and thus by taking to pen and ink, we should soon settle the question.

—Gottfried Wilhelm Leibniz [1677]

AI: computational agents that act intelligently

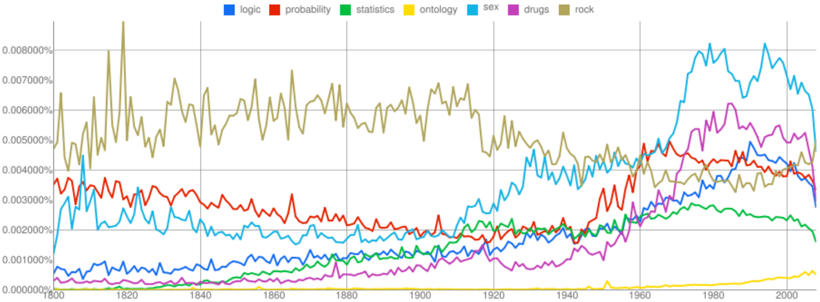


Logic, Probability, Statistics, Ontology over time



From: Google Books Ngram Viewer
(http://ngrams.googlelabs.com/)

Logic, Probability, Statistics, Sex, Drugs, Rock



From: Google Books Ngram Viewer (http://ngrams.googlelabs.com/)

Outline

1. Logic and Probability
   - Relational Probabilistic Models
   - Probabilistic Programming Languages
   - Probabilistic Logic Programs
   - Lifted Inference
2. Semantic Science Overview
   - Ontologies
   - Data
   - Hypotheses and Theories
   - Models
3. Existence and Identity Uncertainty

Why Logic?

Logic provides a **semantics** linking

- the symbols in our language
- the (real or imaginary) world we are trying to characterise

Suppose K represents our knowledge of the world

- If

  $$K \models g$$

  then g must be true of the world.
- If

  $$K \not\models g$$

  there is a model of K in which g is false.

Thus logical consequence seems like the correct notion for prediction.

First-order Predicate Calculus

The world (we want to represent) is made up of individuals (things) and relationships between things.

Classical (first order) logic lets us represent:

- individuals in the world
- relations amongst those individuals
- conjunctions, disjunctions, negations of relations
- quantification over individuals

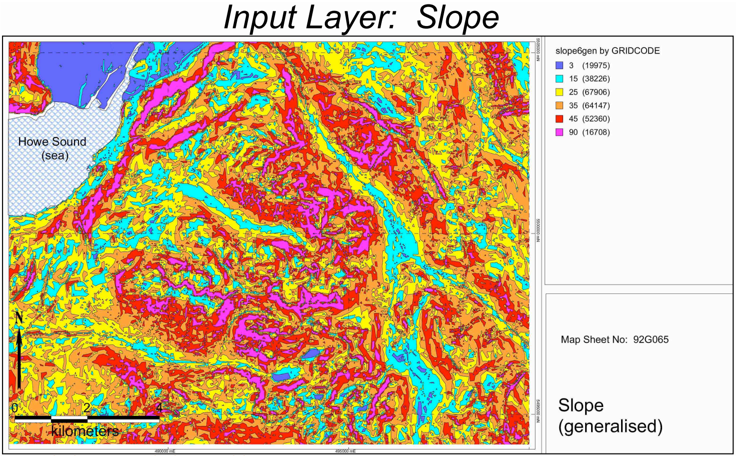
Why Probability?

- There is lots of uncertainty about the world, but agents still need to act.
- Predictions are needed to decide what to do:
  - definitive predictions: you will be run over tomorrow
  - point probabilities: probability you will be run over tomorrow is 0.002
  - probability ranges: you will be run over with probability in range [0.001,0.34]
- Acting is gambling: agents who don't use probabilities will lose to those who do — Dutch books.
- Probabilities can be learned from data.
  Bayes' rule specifies how to combine data and prior knowledge.

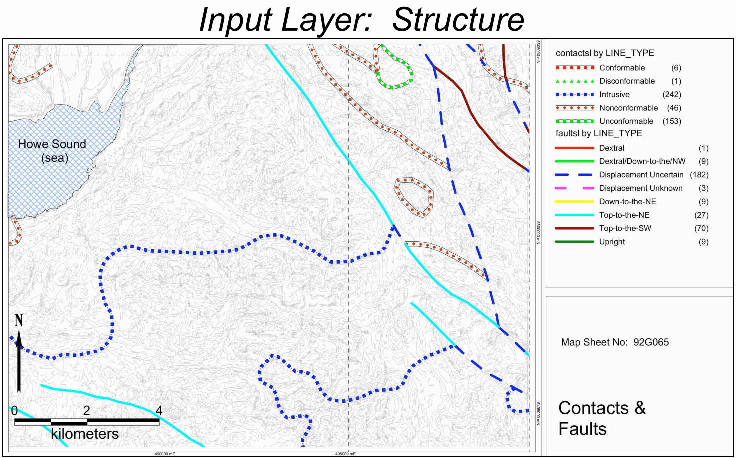
Bayes' Rule

$$P(h|e) = \frac{\overset{\text{Likelihood}}{P(e|h)}\ \ \overset{\text{Prior}}{P(h)}}{\underset{\text{Normalizing constant}}{P(e)}}$$

Example Observation, Geology



Example Observation, Geology



Outline

1. Logic and Probability
   - Relational Probabilistic Models
   - Probabilistic Programming Languages
   - Probabilistic Logic Programs
   - Lifted Inference
2. Semantic Science Overview
   - Ontologies
   - Data
   - Hypotheses and Theories
   - Models
3. Existence and Identity Uncertainty

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:

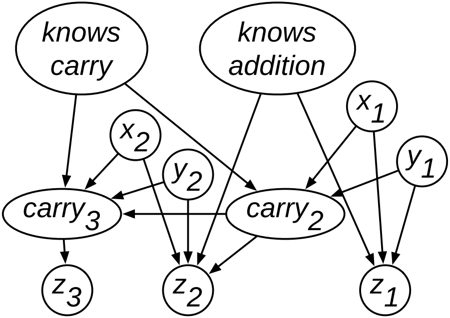
$$
\begin{aligned}
& prop(joe, likes, R) \leftarrow \\
& \qquad prop(R, type, resort) \wedge \\
& \qquad prop(R, near, B) \wedge \\
& \qquad prop(B, type, beach) \wedge \\
& \qquad prop(B, covered\_in, S) \wedge \\
& \qquad prop(S, type, sand).
\end{aligned}
$$

Joe likes resorts that are near sandy beaches.

- But we want probabilistic predictions.

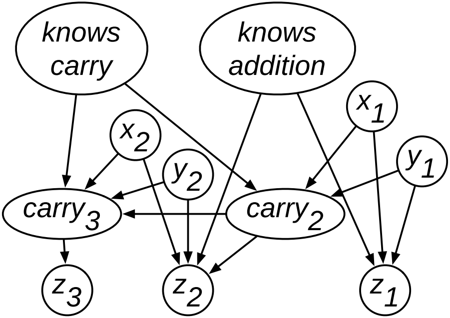
Bayesian Networks

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



Bayesian Networks

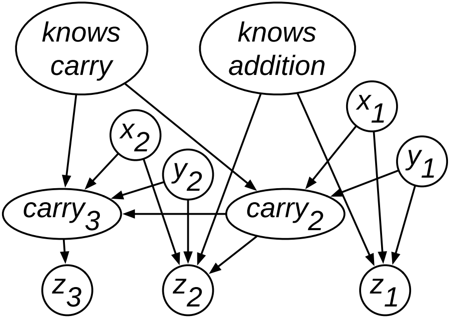
$$\begin{array}{rccc} & & x_2 & x_1 \\ + & & y_2 & y_1 \\ \hline & z_3 & z_2 & z_1 \end{array}$$



What if there were multiple digits

Bayesian Networks

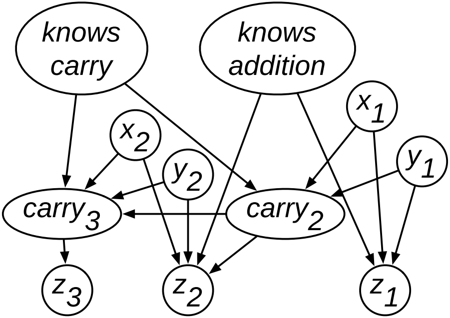
$$
\begin{array}{rrrr}
 & & x_2 & x_1 \\
+ & & y_2 & y_1 \\
\hline
 & z_3 & z_2 & z_1
\end{array}
$$



What if there were multiple digits, problems

Bayesian Networks

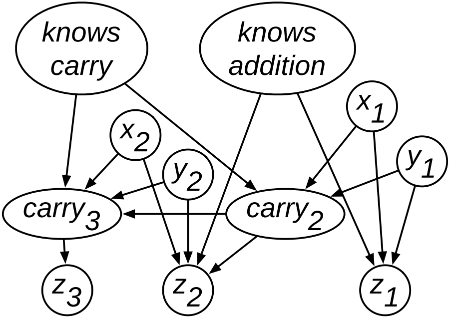
$$
\begin{array}{rccc}
 & & x_2 & x_1 \\
+ & & y_2 & y_1 \\
\hline
 & z_3 & z_2 & z_1
\end{array}
$$



What if there were multiple digits, problems, students

Bayesian Networks

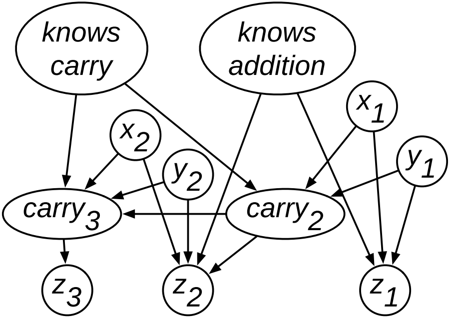
$$\begin{array}{cccc} & & x_2 & x_1 \\ + & & y_2 & y_1 \\ \hline & z_3 & z_2 & z_1 \end{array}$$



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

Bayesian Networks

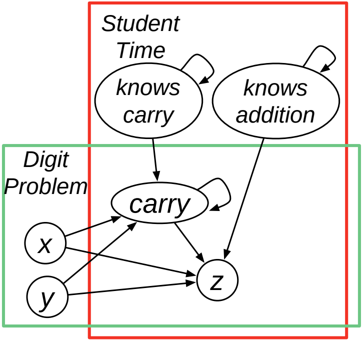
$$\begin{array}{cccc} & & x_2 & x_1 \\ + & & y_2 & y_1 \\ \hline & z_3 & z_2 & z_1 \end{array}$$



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

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



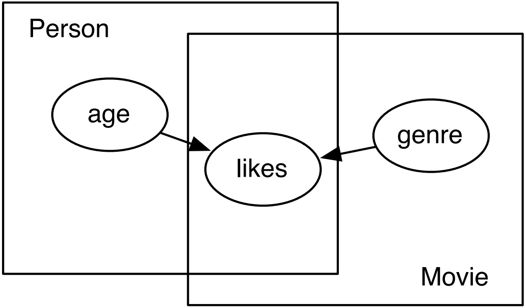
Random Variables: $x(D, P)$, $y(D, P)$, $knowsCarry(S, T)$, $knowsAddition(S, T)$, $carry(D, P, S, T)$, $z(D, P, S, T)$
for each: digit D, problem P, student S, time T

☛ parametrized random variables

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: collaborative filtering

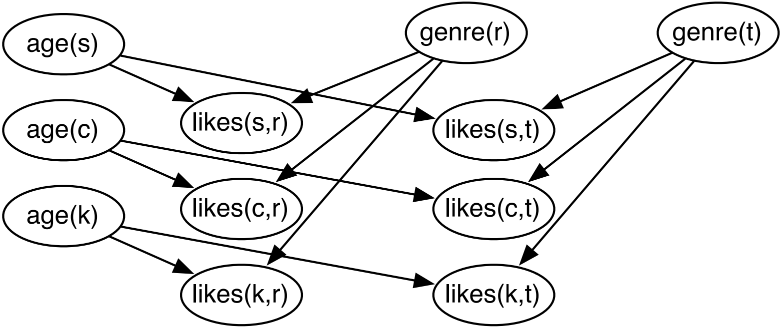


Parametrized random variables: $age(P)$, $likes(P, M)$, $genre(M)$.
If there are 1000 people and 100 movies,
Grounding contains: 100,000 likes + 1,000 age + 100 genre = 101,100 random variables

Example: collaborative filtering

The network means its grounding:

- the population of *Person* is $\{sam, chris, kim\}$
- the population of *Movie* is $\{terminator, rango\}$



Representing Conditional Probabilities

- $P(knows\_addition(X)|bright(X), taught\_addition(X))$
  parameter sharing — individuals share probability parameters.
- $P(happy(X)|friend(X, Y), mean(Y))$
  needs aggregation — $happy(a)$ depends on an unbounded number of parents.
- the carry of one digit depends on carry of the previous digit

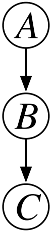
Outline

1. Logic and Probability
   - Relational Probabilistic Models
   - Probabilistic Programming Languages
   - Probabilistic Logic Programs
   - Lifted Inference
2. Semantic Science Overview
   - Ontologies
   - Data
   - Hypotheses and Theories
   - Models
3. Existence and Identity Uncertainty

Probabilistic Programming Languages

- Probabilistic inputs (used in Simula in 1966)
- Conditioning on observations, and querying for distributions
- Inference: more efficient than rejection sampling
- Learning probabilities from data

Representing Bayesian networks



$$\begin{aligned} P(a) &= 0.1 \\ P(b|a) &= 0.8 \\ P(b|\neg a) &= 0.3 \\ P(c|b) &= 0.4 \\ P(c|\neg b) &= 0.75 \end{aligned}$$

$P(a) = 0.1,$
$P(bifa) = 0.8,\ P(bifna) = 0.3,$
$P(cifb) = 0.4,\ P(cifnb) = 0.75.$
$b \iff (a \wedge bifa) \vee (\neg a \wedge bifna)$
$c \iff (b \wedge cifb) \vee (\neg b \wedge cifnbc)$

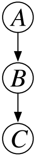
```
begin
   Boolean a,b,c;
   a := draw(0.1);
   if a then
      b := draw(0.8);
   else
      b := draw(0.3);
   if b then
      c := draw(0.4);
   else
      c := draw(0.75);
end
```

Semantics of Probabilistic Programming Languages

"Alternative" for each instance of a probabilistic input possibly encountered in an execution of a program.

- Rejection sampling
- Independent choice: possible world for each assignment of a value for each alternative; program specifies what is true in each world
- Program trace semantics: possible world for each choice encountered in execution path
- Abductive semantics: possible world for each choice needed to infer observations and a value for a query

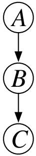
Independent Choice Semantics



$$P(a) = 0.1, P(bifa) = 0.8,\ P(bifna) = 0.3,$$
$$P(cifb) = 0.4,\ P(cifnb) = 0.75.$$
$$b \iff (a \wedge bifa) \vee (\neg a \wedge bifna)$$
$$c \iff (b \wedge cifb) \vee (\neg b \wedge cifnbc)$$

World	*A*	*Bifa*	*Bifna*	*Cifb*	*Cifnb*	Probability
w_0	false	false	false	false	false	$0.9 \cdot 0.2 \cdot 0.7 \cdot 0.6 \cdot 0.25$
w_1	false	false	false	false	true	$0.9 \cdot 0.2 \cdot 0.7 \cdot 0.6 \cdot 0.75$
...						
w_{30}	true	true	true	true	false	$0.1 \cdot 0.8 \cdot 0.3 \cdot 0.4 \cdot 0.75$
w_{31}	true	true	true	true	true	$0.1 \cdot 0.8 \cdot 0.3 \cdot 0.4 \cdot 0.75$

Program Trace Semantics



$$P(a) = 0.1, P(bifa) = 0.8,\ P(bifna) = 0.3,$$
$$P(cifb) = 0.4,\ P(cifnb) = 0.75.$$
$$b \iff (a \wedge bifa) \vee (\neg a \wedge bifna)$$
$$c \iff (b \wedge cifb) \vee (\neg b \wedge cifnbc)$$

World	A	$Bifa$	$Bifna$	$Cifb$	$Cifnb$	Probability
w_0	false	$\perp$	false	$\perp$	false	$0.9 \times 0.7 \times 0.25$
w_1	false	$\perp$	false	$\perp$	true	$0.9 \times 0.7 \times 0.75$
...						
w_7	true	true	$\perp$	false	$\perp$	$0.1 \times 0.8 \times 0.6$
w_8	true	true	$\perp$	true	$\perp$	$0.1 \times 0.8 \times 0.4$

Abductive semantics for computing $P(q|obs)$, only need minimum set of choices needed to infer $obs \wedge q$ or $obs \wedge \neg q$.

Outline

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.
- 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 atomic hypothesis is an atomic formula.
  An alternative is a set of atomic hypotheses.
  $\mathcal{C}$, the choice space is a set of disjoint alternatives.
- $\mathcal{F}$, the facts is an acyclic logic program that gives consequences of choices (can contain negation as failure).
  No atomic hypothesis is the head of a rule.
- 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

- There is a possible world for each selection of one element from each alternative.
- The logic program together with the selected atoms specifies what is true in each possible world.
- The elements of different alternatives are independent.

Meaningless Example: Semantics

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

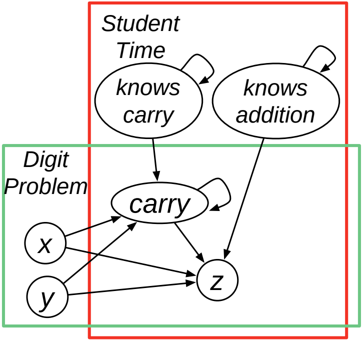
$P_0(c_1) = 0.5 \quad P_0(c_2) = 0.3 \quad P_0(c_3) = 0.2$
$P_0(b_1) = 0.9 \quad P_0(b_2) = 0.1$

$$\begin{array}{llllllll} & & \overbrace{}^{\text{selection}} & & \overbrace{}^{\text{logic program}} & & & \\ w_1 & \models & c_1 & b_1 & f & d & e & P(w_1) = 0.45 \\ w_2 & \models & c_2 & b_1 & \neg f & \neg d & e & P(w_2) = 0.27 \\ w_3 & \models & c_3 & b_1 & \neg f & d & \neg e & P(w_3) = 0.18 \\ w_4 & \models & c_1 & b_2 & \neg f & d & \neg e & P(w_4) = 0.05 \\ w_5 & \models & c_2 & b_2 & \neg f & \neg d & e & P(w_5) = 0.03 \\ w_6 & \models & c_3 & b_2 & f & \neg d & e & P(w_6) = 0.02 \end{array}$$

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

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

☛ parametrized random variables

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:

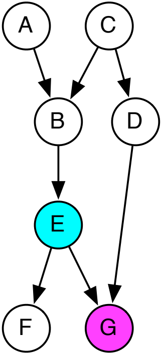
$\forall DPST \{noMistake(D,P,S,T), mistake(D,P,S,T)\}$

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

Outline

1. Logic and Probability
   - Relational Probabilistic Models
   - Probabilistic Programming Languages
   - Probabilistic Logic Programs
   - Lifted Inference
2. Semantic Science Overview
   - Ontologies
   - Data
   - Hypotheses and Theories
   - Models
3. Existence and Identity Uncertainty

Bayesian Network Inference



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

$$P(E \wedge g) = \sum_F \sum_B \sum_C \sum_A \sum_D P(A)P(B|AC)$$

$$P(C)P(D|C)P(E|B)P(F|E)P(g|ED)$$

$$= \left(\sum_F P(F|E)\right)$$

$$\sum_B P(e|B) \sum_C P(C) \left(\sum_A P(A)P(B|AC)\right)$$

$$\left(\sum_D P(D|C)P(g|ED)\right)$$

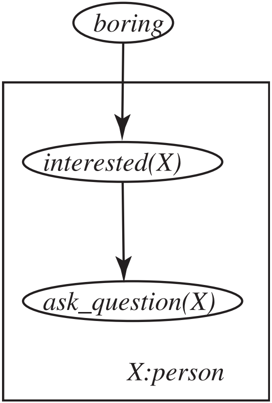
Exchangeability

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

Lifted Inference

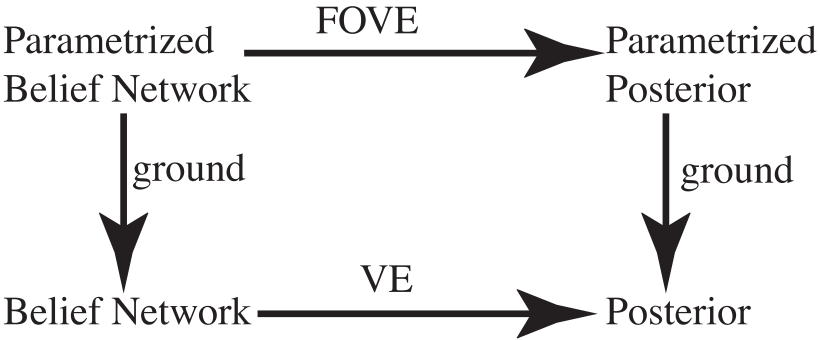
- Idea: treat those individuals about which you have the same information as a block; just count them.
- Use the ideas from lifted theorem proving - no need to ground.
- Potential to be exponentially faster in the number of non-differentialed individuals.
- Relies on knowing the number of individuals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

Example parametrized belief network



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

First-order probabilistic inference



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

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

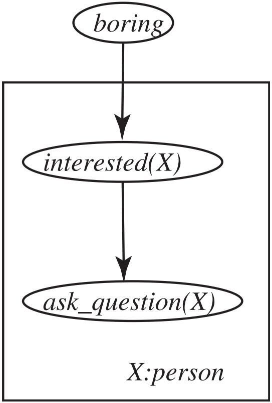
Parametric Factors

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.

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

Removing a parameter when summing



n people

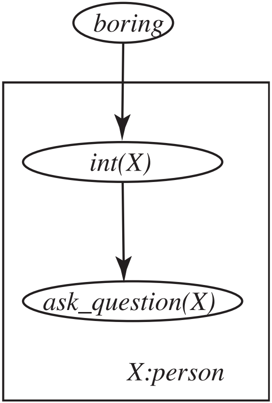
we observe no questions

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

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

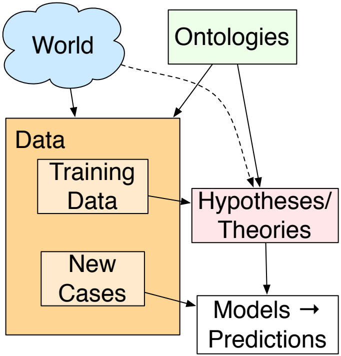
- If a KR system makes a prediction, we should ask: what evidence is there? The system should be able to provide such evidence.
- A knowledge-based system should believe based on evidence. Not all beliefs are equally valid.
- The mechanism that has been developed for judging knowledge is called science.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



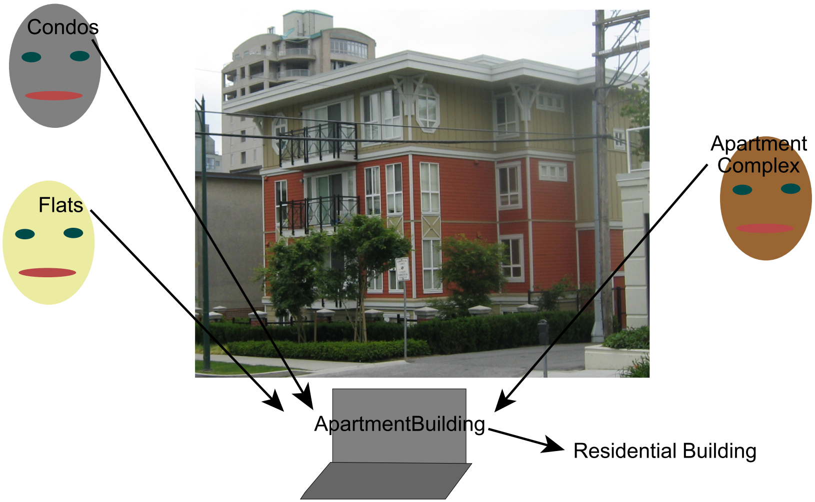
- Ontologies represent the meaning of symbols.
- Data that adheres to ontologies are published.
- Hypotheses that make (probabilistic) predictions on data are published.
- Data used to evaluate hypotheses; the best hypotheses are theories.
- Hypotheses form models for predictions on new cases.
- All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

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 are defined in terms of properties.
- Aristotelean definitions provide the (parametrized) random variables.

Outline

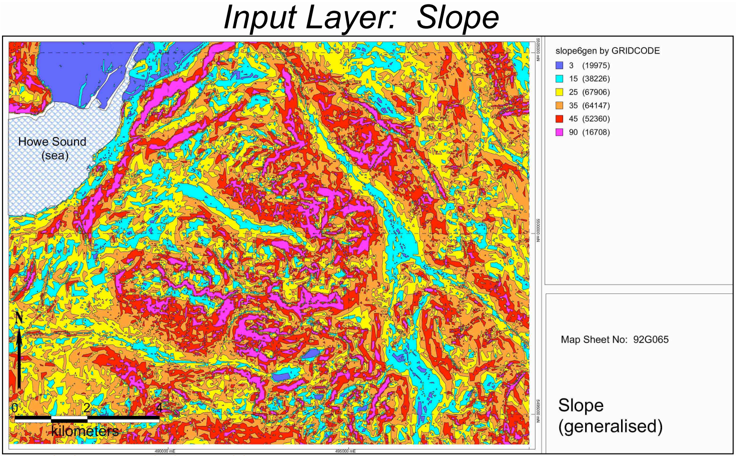
1. Logic and Probability
   - Relational Probabilistic Models
   - Probabilistic Programming Languages
   - Probabilistic Logic Programs
   - Lifted Inference
2. Semantic Science Overview
   - Ontologies
   - Data
   - Hypotheses and Theories
   - Models
3. Existence and Identity Uncertainty

Data

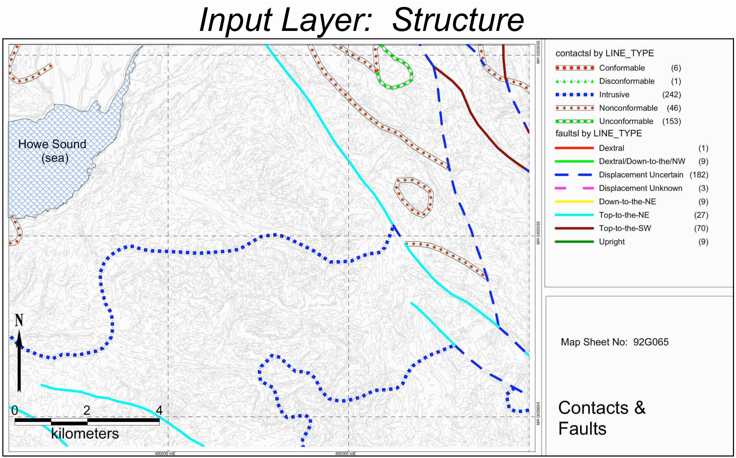
Real data is messy!

- 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



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

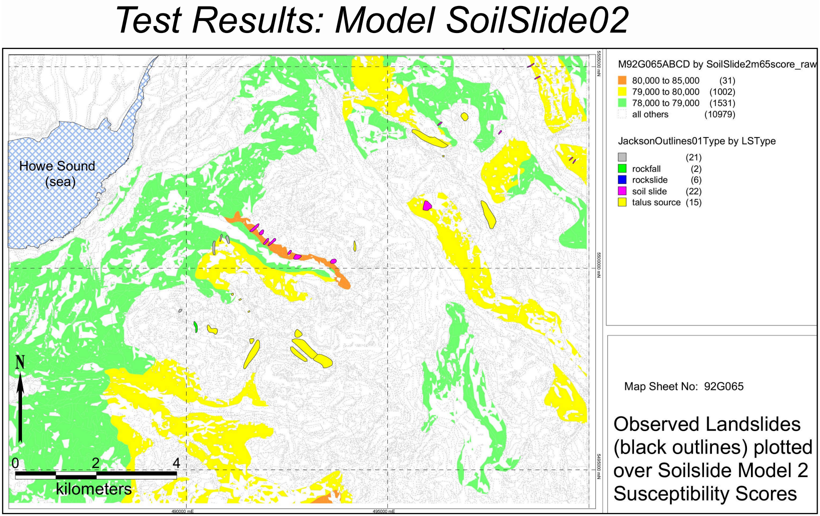
1. Logic and Probability
    - Relational Probabilistic Models
    - Probabilistic Programming Languages
    - Probabilistic Logic Programs
    - Lifted Inference
2. Semantic Science Overview
    - Ontologies
    - Data
    - Hypotheses and Theories
    - Models
3. Existence and Identity Uncertainty

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

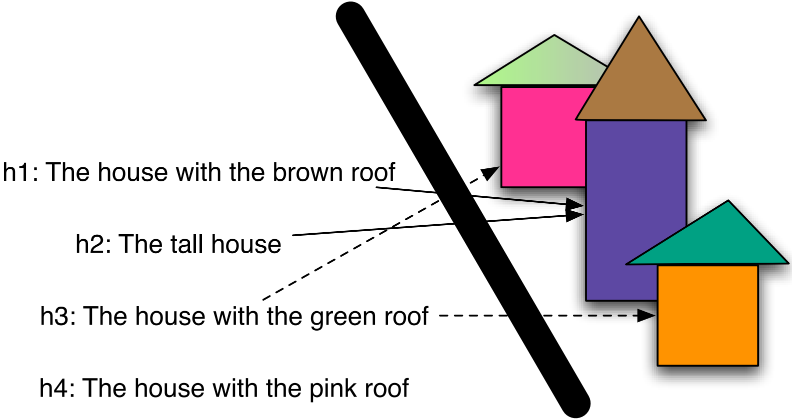
Dynamics of Semantic Science

- 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.

Outline

1. Logic and Probability
   - Relational Probabilistic Models
   - Probabilistic Programming Languages
   - Probabilistic Logic Programs
   - Lifted Inference
2. Semantic Science Overview
   - Ontologies
   - Data
   - Hypotheses and Theories
   - Models
3. Existence and Identity Uncertainty

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?

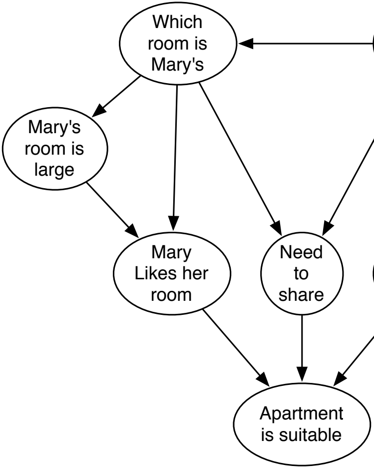
Role assignments

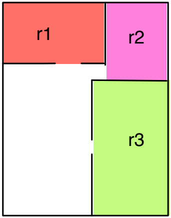
Hypothesis about what apartment Mary would like.

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

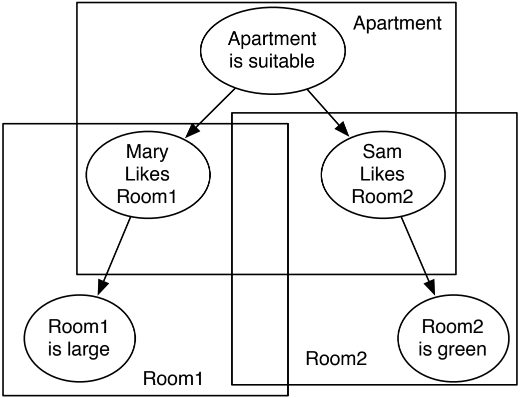
BN Representation

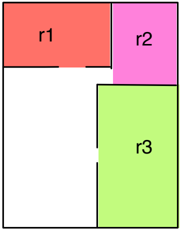




How can we condition on the observation of the apartment?

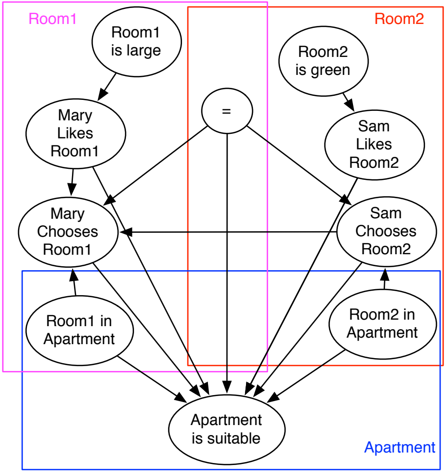
Naive Bayes representation

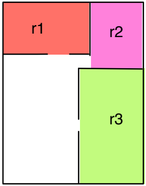




How do we specify that Mary chooses a room?
What about the case where they (have to) share?

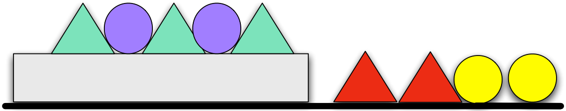
Causal representation





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

Observation Protocols

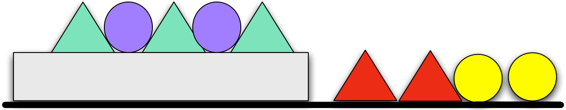


Observe a triangle and a circle touching. What is the probability the triangle is green?

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

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

Protocol for Observing



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

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

Conclusion

- To decide what to do an agent should take into account its uncertainty and it preferences (utility).
- The field of "statistical relational AI" looks at how to combine first-order logic and probabilistic reasoning.
- We need both (prior) knowledge and data to make predictions needed for action.

Challenges

- Knowledge representations that are heuristically and epistemologically adequate and take into account all data that can be obtained.
- Combine representations with ontologies to interoperate with heterogenous data sets and predictions made by various hypotheses developed by different people.

Bayes' Rule

$$P(h|e) = \frac{\overbrace{P(e|h)}^{\text{Likelihood}}\ \ \overbrace{P(h)}^{\text{Prior}}}{\underbrace{P(e)}_{\text{Normalizing constant}}}$$

AI: computational agents that act intelligently