Logic, Probability and Computation: Statistical Relational AI and Beyond

David Poole

Department of Computer Science,
University of British Columbia

October 2014
AI: computational agents that act intelligently

What should an agent do?

Logic
Probability
Ontologies
Knowledge Representation
Learning
Relations
Preferences/Utilities
Decision Theory
Inference
Knowledge Aquisition
Perceiving
Modelling
Diagnosis
Acting

Tasks

Inputs

Ontologies
Prior Knowledge
Observations
Data
Relations
Hypotheses
Preferences/Utilities
Abilities
Dynamical Systems

Foundations

Logic
Probability
Decision Theory
Statistics
Computation
Game theory
Knowledge Representation
Outline

1. Logic and Probability
   - Relational Probabilistic Models
   - Probabilistic Logic Programs

2. Lifted Inference

3. Undirected models, Directed models, and Weighted Formulae

4. 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

\[
P(h|e) = \frac{P(e|h) \cdot P(h)}{P(e)}
\]

Likelihood Prior

Normalizing constant

What if \( e \) is an electronic health record?
Bayes’ Rule

\[ P(h|e) = \frac{P(\text{elh}) \cdot P(h)}{P(e)} \]

Likelihood Prior
Normalizing constant

What if \( e \) is an electronic health record?
What if \( e \) is all the electronic health records?
Bayes’ Rule

\[
P(h|e) = \frac{P(e|h) P(h)}{P(e)}
\]

What if \( e \) is an electronic health record? What if \( e \) is all the electronic health records? What if \( e \) is a description of everything known about the geology of Earth?
Example Observation, Geology

**Input Layer: Slope**

Map Sheet No: 92G065

Slope (generalised)
Example Observation, Geology

Input Layer: Structure

Map Sheet No: 92G065

Contacts & Faults

Howe Sound (sea)
Outline

1. Logic and Probability
   - Relational Probabilistic Models
   - Probabilistic Logic Programs

2. Lifted Inference

3. Undirected models, Directed models, and Weighted Formulae

4. 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?

<table>
<thead>
<tr>
<th>Individual</th>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>joe</td>
<td>likes</td>
<td>resort_14</td>
</tr>
<tr>
<td>joe</td>
<td>dislikes</td>
<td>resort_35</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>resort_14</td>
<td>type</td>
<td>resort</td>
</tr>
<tr>
<td>resort_14</td>
<td>near</td>
<td>beach_18</td>
</tr>
<tr>
<td>beach_18</td>
<td>type</td>
<td>beach</td>
</tr>
<tr>
<td>beach_18</td>
<td>covered_in</td>
<td>ws</td>
</tr>
<tr>
<td>ws</td>
<td>type</td>
<td>sand</td>
</tr>
<tr>
<td>ws</td>
<td>color</td>
<td>white</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Example: trading agent

Possible theory that could be learned:

\[
\text{prop(joe, likes, R) } \leftarrow \text{prop(R, type, resort)} \land \\
\text{prop(R, near, B)} \land \\
\text{prop(B, type, beach)} \land \\
\text{prop(B, covered_in, S)} \land \\
\text{prop(S, type, sand)}.}
\]

Joe likes resorts that are near sandy beaches.

- But we want probabilistic predictions.
### Example: Predicting Relations

<table>
<thead>
<tr>
<th>Student</th>
<th>Course</th>
<th>Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_1$</td>
<td>$c_1$</td>
<td>$A$</td>
</tr>
<tr>
<td>$s_2$</td>
<td>$c_1$</td>
<td>$C$</td>
</tr>
<tr>
<td>$s_1$</td>
<td>$c_2$</td>
<td>$B$</td>
</tr>
<tr>
<td>$s_2$</td>
<td>$c_3$</td>
<td>$B$</td>
</tr>
<tr>
<td>$s_3$</td>
<td>$c_2$</td>
<td>$B$</td>
</tr>
<tr>
<td>$s_4$</td>
<td>$c_3$</td>
<td>$B$</td>
</tr>
<tr>
<td>$s_3$</td>
<td>$c_4$</td>
<td>?</td>
</tr>
<tr>
<td>$s_4$</td>
<td>$c_4$</td>
<td>?</td>
</tr>
</tbody>
</table>

- Students $s_3$ and $s_4$ have the same averages, on courses with the same averages.
- Which student would you expect to better?
From Relations to Belief Networks

```
Gr(s1, c1)
Gr(s2, c1)
Gr(s1, c2)
Gr(s2, c3)
Gr(s3, c2)
Gr(s3, c4)
Gr(s4, c3)
Gr(s4, c4)
```

$$I(s_1) \quad D(c_1)$$
$$I(s_2) \quad Gr(s_1, c_1) \quad Gr(s_2, c_1)$$
$$I(s_2) \quad Gr(s_1, c_2) \quad D(c_2)$$
$$I(s_3) \quad Gr(s_2, c_3) \quad D(c_2)$$
$$I(s_3) \quad Gr(s_3, c_2) \quad D(c_3)$$
$$I(s_4) \quad Gr(s_3, c_4) \quad D(c_4)$$
$$I(s_4) \quad Gr(s_4, c_3) \quad D(c_4)$$
From Relations to Belief Networks

<table>
<thead>
<tr>
<th>$I(S)$</th>
<th>$D(C)$</th>
<th>$Gr(S, C)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>true</td>
<td>true</td>
<td>0.5</td>
</tr>
<tr>
<td>true</td>
<td>false</td>
<td>0.9</td>
</tr>
<tr>
<td>false</td>
<td>true</td>
<td>0.01</td>
</tr>
<tr>
<td>false</td>
<td>false</td>
<td>0.1</td>
</tr>
</tbody>
</table>

$P(I(S)) = 0.5$
$P(D(C)) = 0.5$

“parameter sharing”
Example: Predicting Relations
Plate Notation

- $S$, $C$ logical variable representing students, courses
- the set of individuals of a type is called a population
- $I(S)$, $Gr(S, C)$, $D(C)$ are parametrized random variables
Plate Notation

- \( S, C \) logical variables representing students, courses
- The set of individuals of a type is called a population
- \( I(S), Gr(S, C), D(C) \) are parametrized random variables

Grounding:
- For every student \( s \), there is a random variable \( I(s) \)
- For every course \( c \), there is a random variable \( D(c) \)
- For every \( s, c \) pair there is a random variable \( Gr(s, c) \)
- All instances share the same structure and parameters
If there were 1000 students and 100 courses:
Grounding contains
- 1000 $I(s)$ variables
- 100 $D(C)$ variables
- 100000 $Gr(s, c)$ variables

total: 101100 variables

Numbers to be specified to define the probabilities:
1 for $I(s)$, 1 for $D(C)$, 8 for $Gr(S, C) = 10$ parameters.
Bayesian Networks

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

How can we build a model before we know the individuals?

David Poole

Logic, Probability and Computation
Bayesian Networks

What if there were multiple digits
Bayesian Networks

What if there were multiple digits, problems
Bayesian Networks

What if there were multiple digits, problems, students
Bayesian Networks

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

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

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

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

parametrized random variables
Relational Probabilistic Models

Often we want random variables for combinations of individual in populations

- build a probabilistic model before knowing the individuals
- learn the model for one set of individuals
- apply the model to new individuals
- allow complex relationships between individuals
Before we know anything about individuals, they are indistinguishable, and so should be treated identically.
Representing Conditional Probabilities

- \( P(\text{grade}(S, C) \mid \text{intelligent}(S), \text{difficult}(C)) \) — parameter sharing — individuals share probability parameters.

- \( P(\text{happy}(X) \mid \text{friend}(X, Y), \text{mean}(Y)) \) — needs aggregation — \text{happy}(a) depends on an unbounded number of parents.

- There can be more structure about the individuals
  - the carry of one digit depends on carry of the previous digit
  - probability that two authors collaborate depends on whether they have a paper authored together
Example: Aggregation

\[
\begin{align*}
\text{x} & \quad \text{Has\_gun(x)} \\
\text{Has\_motive(x,y)} & \quad \text{Has\_opportunity(x,y)} \\
\text{Shot(x,y)} & \\
\text{Someone\_shot(y)} & \quad \text{y}
\end{align*}
\]
Example Plate Notation for Learning Parameters

- $T$ is a logical variable representing tosses of a thumb tack.
- $H(t)$ is a Boolean variable that is true if toss $t$ is heads.
- $\theta$ is a random variable representing the probability of heads.
- Range of $\theta$ is $\{0.0, 0.01, 0.02, \ldots, 0.99, 1.0\}$ or interval $[0, 1]$.
- $P(H(t_i) = \text{true} \mid \theta = p) = p$
- $H(t_i)$ is independent of $H(t_j)$ (for $i \neq j$) given $\theta$: i.i.d. or independent and identically distributed.
Outline

1 Logic and Probability
   - Relational Probabilistic Models
   - Probabilistic Logic Programs

2 Lifted Inference

3 Undirected models, Directed models, and Weighted Formulae

4 Existence and Identity Uncertainty
Independent Choice Logic (ICL)

- A language for relational probabilistic models.
- **Idea**: combine logic and probability, where all uncertainty is 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 $\rightarrow$ probabilistic Horn abduction (IJCAI-91); richer language (negation as failure + choices by other agents $\rightarrow$ 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

\[ C = \{\{c_1, c_2, c_3\}, \{b_1, b_2\}\} \]

\[ \mathcal{F} = \{ f \leftarrow c_1 \land b_1, \quad f \leftarrow c_3 \land b_2, \\
\quad d \leftarrow c_1, \quad d \leftarrow \neg c_2 \land b_1, \\
\quad e \leftarrow f, \quad e \leftarrow \neg d\} \]

\[ P_0(c_1) = 0.5 \quad P_0(c_2) = 0.3 \quad P_0(c_3) = 0.2 \]

\[ P_0(b_1) = 0.9 \quad P_0(b_2) = 0.1 \]
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

\[ \mathcal{F} = \{ f \leftarrow c_1 \land b_1, \quad f \leftarrow c_3 \land b_2, \quad d \leftarrow c_1, \quad d \leftarrow \sim c_2 \land b_1, \quad e \leftarrow f, \quad e \leftarrow \sim d \} \]

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

Selection | Logic Program | Probability
---|---|---
\( w_1 \models c_1 \quad b_1 \quad f \quad d \quad e \) | \( P(w_1) = 0.45 \)
\( w_2 \models c_2 \quad b_1 \quad \sim f \quad \sim d \quad e \) | \( P(w_2) = 0.27 \)
\( w_3 \models c_3 \quad b_1 \quad \sim f \quad d \quad \sim e \) | \( P(w_3) = 0.18 \)
\( w_4 \models c_1 \quad b_2 \quad \sim f \quad d \quad \sim e \) | \( P(w_4) = 0.05 \)
\( w_5 \models c_2 \quad b_2 \quad \sim f \quad \sim d \quad e \) | \( P(w_5) = 0.03 \)
\( w_6 \models c_3 \quad b_2 \quad f \quad \sim d \quad e \) | \( P(w_6) = 0.02 \)

\[ P(e) = 0.45 + 0.27 + 0.03 + 0.02 = 0.77 \]
Belief Networks, Decision trees and ICL rules

- There is a local mapping from belief networks into ICL.

\[
\begin{align*}
\text{prob } \text{ta} &: 0.02. \\
\text{prob } \text{fire} &: 0.01. \\
\text{alarm} &\leftarrow \text{ta} \land \text{fire} \land \text{atf}. \\
\text{alarm} &\leftarrow \sim \text{ta} \land \text{fire} \land \text{antf}. \\
\text{alarm} &\leftarrow \text{ta} \land \sim \text{fire} \land \text{atnf}. \\
\text{alarm} &\leftarrow \sim \text{ta} \land \sim \text{fire} \land \text{antnf}. \\
\text{prob } \text{atf} &: 0.5. \\
\text{prob } \text{antf} &: 0.99. \\
\text{prob } \text{atnf} &: 0.85. \\
\text{prob } \text{antnf} &: 0.0001. \\
\text{smoke} &\leftarrow \text{fire} \land \text{sf}. \\
\text{smoke} &\leftarrow \sim \text{fire} \land \text{snf}. \\
\text{prob } \text{sf} &: 0.9. \\
\text{prob } \text{snf} &: 0.01.
\end{align*}
\]
Belief Networks, Decision trees and ICL rules

- Rules can represent decision tree with probabilities:

\[
P(e|A,B,C,D) = P_0(h_1) = 0.7
\]
\[
e \leftarrow a \land b \land h_1.
\]
\[
P(e|A,B,C,D) = P_0(h_2) = 0.2
\]
\[
e \leftarrow a \land \neg b \land h_2.
\]
\[
P(e|A,B,C,D) = P_0(h_3) = 0.9
\]
\[
e \leftarrow \neg a \land c \land d \land h_3.
\]
\[
P(e|A,B,C,D) = P_0(h_4) = 0.5
\]
\[
e \leftarrow \neg a \land c \land \neg d \land h_4.
\]
\[
P(e|A,B,C,D) = P_0(h_5) = 0.3
\]

\[
e \leftarrow \neg a \land \neg c \land h_5.
\]
Predicting Grades

prob \( int(S) \): 0.5.
prob \( \text{diff}(C) \): 0.5.
\( gr(S, C, G) \) \( \leftarrow \) \( int(S) \) \( \wedge \) \( \text{diff}(C) \) \( \wedge \) \( \text{idg}(S, C, G) \).
prob \( \text{idg}(S, C, a) \): 0.5, \( \text{idg}(S, C, b) \): 0.4, \( \text{idg}(S, C, c) \): 0.1.
\( gr(S, C, G) \) \( \leftarrow \) \( int(S) \) \( \wedge \) \( \sim \) \( \text{diff}(C) \) \( \wedge \) \( \text{idg}(S, C, G) \).
prob \( \text{idg}(S, C, a) \): 0.9, \( \text{idg}(S, C, b) \): 0.09, \( \text{idg}(S, C, c) \): 0.01.
\( gr(S, C, G) \) \( \leftarrow \) \( \sim \) \( int(S) \) \( \wedge \) \( \text{diff}(C) \) \( \wedge \) \( \text{nidg}(S, C, G) \).
prob \( \text{nidg}(S, C, a) \): 0.01, \( \text{nidg}(S, C, b) \): 0.09, \( \text{nidg}(S, C, c) \): 0.9.
\( gr(S, C, G) \) \( \leftarrow \) \( \sim \) \( int(S) \) \( \wedge \) \( \sim \) \( \text{diff}(C) \) \( \wedge \) \( \text{nindg}(S, C, G) \).
prob \( \text{nindg}(S, C, a) \): 0.1, \( \text{nindg}(S, C, b) \): 0.4, \( \text{nindg}(S, C, c) \): 0.5.
Multi-digit addition with parametrized BNs / plates

\[
\begin{align*}
  & x_j x_1 \\
+ & y_j y_1 \\
\hline
  & z_j z_1
\end{align*}
\]

Random Variables: \( x(D, P), y(D, P), \text{knowsCarry}(S, T), \text{knowsAddition}(S, T), \text{carry}(D, P, S, T), z(D, P, S, T) \)

for each: digit \( D \), problem \( P \), student \( S \), time \( T \)

parametrized random variables
ICL rules for multi-digit addition

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

\[ z(D, P, S, T) = V \leftarrow \]
\[ \text{knowsAddition}(S, T) \wedge \]
\[ \text{mistake}(D, P, S, T) \wedge \]
\[ \text{selectDig}(D, P, S, T) = V. \]

\[ z(D, P, S, T) = V \leftarrow \]
\[ \neg \text{knowsAddition}(S, T) \wedge \]
\[ \text{selectDig}(D, P, S, T) = V. \]

Alternatives:
\[ \forall DPST \{ \text{noMistake}(D, P, S, T), \text{mistake}(D, P, S, T) \} \]
\[ \forall DPST \{ \text{selectDig}(D, P, S, T) = V \mid V \in \{0..9\} \} \]
Outline

1. Logic and Probability
   - Relational Probabilistic Models
   - Probabilistic Logic Programs

2. Lifted Inference

3. Undirected models, Directed models, and Weighted Formulae

4. Existence and Identity Uncertainty
Bayesian Network Inference

\[
P(E \mid g) = \frac{P(E \land g)}{p(g)}
\]

\[
P(E \land g) = \sum_F \sum_B \sum_C \sum_A \sum_D P(A) P(B \mid AC) \cdot P(C) P(D \mid C) P(E \mid B) P(F \mid E) P(g \mid ED)
\]

\[
= \left( \sum_F P(F \mid E) \right)
\]

\[
\sum_B P(e \mid B) \sum_C P(C) \left( \sum_A P(A) P(B \mid AC) \left( \sum_D P(D \mid C) P(g \mid ED) \right) \right)
\]
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).
First-order probabilistic inference

Parametrized Belief Network \(\xrightarrow{\text{FOVE}}\) Parametrized Posterior

Belief Network \(\xrightarrow{\text{ground}}\) Posterior

Belief Network \(\xrightarrow{\text{VE}}\) Posterior
In 1965, Robinson showed how unification allows many ground steps with one step:

\[
\begin{align*}
  f(X, Z) \lor p(X, a) & \sim p(b, Y) \lor g(Y, W) \\
  f(b, Z) \lor g(a, W)
\end{align*}
\]

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:

\[
\begin{align*}
[f(X, Z), p(X, a)] & \times [p(b, Y), g(Y, W)] \\
[f(b, Z), p(b, a), g(a, W)] & \\
\end{align*}
\]

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

\[
\begin{align*}
[f(b, Z), p(b, a)] \\
[f(X, Z), p(X, a)] \text{ with constraint } X \neq b,
\end{align*}
\]
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.

$\langle \{ X \neq sue \}, \{ interested(X), boring \}, \rangle$

<table>
<thead>
<tr>
<th>interested</th>
<th>boring</th>
<th>Val</th>
</tr>
</thead>
<tbody>
<tr>
<td>yes</td>
<td>yes</td>
<td>0.001</td>
</tr>
<tr>
<td>yes</td>
<td>no</td>
<td>0.01</td>
</tr>
</tbody>
</table>

...
Removing a parameter when summing

$n$ people we observe no questions

Eliminate \textit{interested}:

\[ \langle \{ \}, \{ \text{boring, interested}(X) \}, t_1 \rangle \]

\[ \langle \{ \}, \{ \text{interested}(X) \}, t_2 \rangle \]

\[ \downarrow \]

\[ \langle \{ \}, \{ \text{boring} \}, (t_1 \times t_2)^n \rangle \]

\( (t_1 \times t_2)^n \) is computed point-wise; we can compute it in time \( O(\log n) \).
Counting Elimination

Eliminate *boring*:

VE: factor on \{\text{int}(p_1), \ldots, \text{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:**

<table>
<thead>
<tr>
<th>#interested</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>(v_0)</td>
</tr>
<tr>
<td>1</td>
<td>(v_1)</td>
</tr>
<tr>
<td>\ldots</td>
<td>\ldots</td>
</tr>
<tr>
<td>(n)</td>
<td>(v_n)</td>
</tr>
</tbody>
</table>

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

[de Salvo Braz et al. 2007] and [Milch et al. 08]
Potential of Lifted Inference

- Reduce complexity:
  
  \[ \text{polynomial} \rightarrow \text{logarithmic} \]

  \[ \text{exponential} \rightarrow \text{polynomial} \]

- We can now do lifting for unary relations, but we know we can’t do all binary relations [Guy Van den Broeck, 2013]

- An active research area.
Outline

1. Logic and Probability
   - Relational Probabilistic Models
   - Probabilistic Logic Programs

2. Lifted Inference

3. Undirected models, Directed models, and Weighted Formulae

4. Existence and Identity Uncertainty
Logistic Regression

Logistic Regression, write $R(a_i)$ as $R_i$:

$$P(Q|R_1, \ldots, R_n) = \frac{1}{1 + e^{w_0 + w_1 R_1 + \cdots + w_n R_n}}$$

If all of the $R_i$ are exchangeable $w_1, \ldots, w_n$ must all be the same:

$$P(Q|R_1, \ldots, R_n) = \frac{1}{1 + e^{w_0 + w_1 (R_1 + \cdots + R_n)}}$$
Logistic Regression

Logistic Regression, write $R(a_i)$ as $R_i$:

$$P(Q|R_1, \ldots, R_n) = \frac{1}{1 + e^{w_0 + w_1 R_1 + \cdots + w_n R_n}}$$

If all of the $R_i$ are exchangeable $w_1, \ldots, w_n$ must all be the same:

$$P(Q|R_1, \ldots, R_n) = \frac{1}{1 + e^{w_0 + w_1 (R_1 + \cdots + R_n)}}$$

If we learn the parameters for $n = 10$ the prediction for $n = 20$ depends on how values $R_i$ are represented numerically:

- If $True = 1$ and $False = 0$ then $P(Q|R_1, \ldots, R_n)$ depends on the number of $R_i$ that are true.
- If $True = 1$ and $False = -1$ then $P(Q|R_1, \ldots, R_n)$ depends on how many more of $R_i$ are true than false.
- If $True = 0$ and $False = -1$ then $P(Q|R_1, \ldots, R_n)$ depends on the number of $R_i$ that are false.
Directed and Undirected models

- **Weighted formula (WF):** $\langle L, F, w \rangle$
  - $L$ is a set of logical variables,
  - $F$ is a logical formula: $\{\text{free logical variables in } F\} \subseteq L$
  - $w$ is a real-valued weight.

- **Instances** of weighted formula obtained by assigning individuals to variables in $L$. 

---

[Note: The content is a continuation of the previous page, discussing directed and undirected models. The page number and author information are present.]
Directed and Undirected models

- **Weighted formula (WF):** $\langle L, F, w \rangle$
  - $L$ is a set of logical variables,
  - $F$ is a logical formula: $\{\text{free logical variables in } F\} \subseteq L$
  - $w$ is a real-valued weight.

- **Instances** of weighted formmule obtained by assigning individuals to variables in $L$.

- A **world** is an assignment of a value to each ground instance of each atom.

- **Markov logic network (MLN):** “undirected model” weighted formulae define measures on worlds.
Directed and Undirected models

- **Weighted formula (WF):** \( \langle L, F, w \rangle \)
  - \( L \) is a set of logical variables,
  - \( F \) is a logical formula: \( \{ \text{free logical variables in } F \} \subseteq L \)
  - \( w \) is a real-valued weight.

- **Instances** of weighted formula obtained by assigning individuals to variables in \( L \).

- **A world** is an assignment of a value to each ground instance of each atom.

- **Markov logic network (MLN):** “undirected model” weighted formulae define measures on worlds.

- **Relational logistic regression (RLR):** “directed model” weighted formulae define conditional probabilities.
Example

Weighted formulae:

\[ \langle \{x\}, \text{funFor}(x), -5 \rangle \]
\[ \langle \{x, y\}, \text{funFor}(x) \land \text{knows}(x, y) \land \text{social}(y), 10 \rangle \]

If \( obs \) includes observations for all \( \text{knows}(x, y) \) and \( \text{social}(y) \):

\[ P(\text{funFor}(x) \mid obs) = \text{sigmoid}(-5 + 10n_T) \]

\( n_T \) is the number of individuals \( y \) for which \( \text{knows}(x, y) \land \text{social}(y) \) is True in \( obs \).

\[ \text{sigmoid}(x) = \frac{1}{1 + e^{-x}} \]
Abstract Example

\[
\langle \{\}, q, \alpha_0 \rangle \\
\langle \{x\}, q \land \neg r(x), \alpha_1 \rangle \\
\langle \{x\}, q \land r(x), \alpha_2 \rangle \\
\langle \{x\}, r(x), \alpha_3 \rangle
\]

If \( r(x) \) for every individual \( x \) is observed:

\[
P(q \mid obs) = \text{sigmoid}(\alpha_0 + n_F \alpha_1 + n_T \alpha_2)
\]

\( n_T \) is number of individuals for which \( r(x) \) is true
\( n_F \) is number of individuals for which \( r(x) \) is false

\[
\text{sigmoid}(x) = \frac{1}{1 + e^{-x}}
\]
Three Elementary Models

(a) Naïve Bayes
(b) (Relational) Logistic Regression
(c) Markov network
Independence Assumptions

- Naïve Bayes and Markov network: \( R(x) \) and \( R(y) \) (for \( x \neq y \))
  - are independent given \( Q \)
  - are dependent not given \( Q \).
- Directed model with aggregation: \( R(x) \) and \( R(y) \) (for \( x \neq y \))
  - are dependent given \( Q \),
  - are independent not given \( Q \).
What happens as Population size \( n \) Changes: Simplest case

\[
\begin{align*}
\langle \emptyset, q, \alpha_0 \rangle \\
\langle \{x\}, q \land \neg r(x), \alpha_1 \rangle \\
\langle \{x\}, q \land r(x), \alpha_2 \rangle \\
\langle \{x\}, r(x), \alpha_3 \rangle \\
\end{align*}
\]

\[
P_{MLN}(q \mid n) = \text{sigmoid}(\alpha_0 + n \log(e^{\alpha_2} + e^{\alpha_1 - \alpha_3}))
\]

\[
P_{RLR}(q \mid n) = \sum_{i=0}^{n} \binom{n}{i} \text{sigmoid}(\alpha_0 + i\alpha_1 + (n-i)\alpha_2)(1-p_r)^i p_r^{n-i}
\]

\[
P_{MF}(q \mid n) = \text{sigmoid}(\alpha_0 + np_r\alpha_1 + n(1 - p_r)\alpha_2)
\]
Population Growth: $P(q \mid n)$
Population Growths: $P_{RLR}(q | n)$

Whereas this MLN is a sigmoid of $n$, RLR needn’t be monotonic:
Outline

1. Logic and Probability
   - Relational Probabilistic Models
   - Probabilistic Logic Programs

2. Lifted Inference

3. Undirected models, Directed models, and Weighted Formulae

4. Existence and Identity Uncertainty
Correspondence Problem

Symbols

h1: The house with the brown roof
h2: The tall house
h3: The house with the green roof
h4: The house with the pink roof

c symbols and i individuals $\rightarrow c^{i+1}$ correspondences
Clarity Principle

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

- What if an individual doesn’t exist?
  
  - `house(h4) ∧ roof_colour(h4, pink) ∧ ¬exists(h4)`
Clarity Principle

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

- What if an individual doesn’t exist?
  - \( \text{house}(h4) \land \text{roof \_colour}(h4, \text{pink}) \land \neg \text{exists}(h4) \)

- 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?
  - \(\text{house}(h4) \land \text{roof-colour}(h4, \text{pink}) \land \neg \exists \text{exists}(h4)\)

- 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.
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
Bayesian Network Representation

Which room is Mary's
Mary's room is large
Mary Likes her room

Which room is Sam's
Sam's room is green
Sam likes her room

Need to share

Apartment is suitable

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?
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

- **false alarm**

- **observe blip**
  - **another blip**
  - **third blip**

- **false alarm**
  - **same plane**
  - **another plane**

- **true alarm**
  - **plane**
  - **false alarm**

- **false alarm**
  - **same plane**
  - **another plane**

- **true alarm**
  - **same plane**
  - **another plane**

- **false alarm**
  - **same plane**
  - **another plane**

- **true alarm**
  - **second plane**

- **false alarm**

- **true alarm**

- **true alarm**

- **true alarm**

- **false alarm**

- **true alarm**

- **true alarm**
First-order Semantic Trees

Split on quantified first-order formulae:

\[
\exists x: \tau(x)
\]

• 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. \( \exists a: \text{apartment}(a) \)
   - f \( \exists r_1: \text{bedroom}(r_1) \land \text{in}(r_1, a) \)
   - t

2. \( \exists r_2: \text{room}(r_2) \land \text{in}(r_2, a) \land \text{green}(r_2) \)
   - f
   - t

3. \( \exists r_3: \text{room}(r_3) \land \text{in}(r_3, a) \land \text{green}(r_3) \)
   - f
   - t

4. \( \exists r_4: \text{room}(r_4) \land \text{in}(r_4, a) \land \text{green}(r_4) \)
   - f
   - t

- 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
Distributions over number

\[\exists c_1: \text{chair}(c_1)\]

\[\exists c_2: \text{chair}(c_2) \land c_1 \neq c_2\]

\[\exists c_3: \text{chair}(c_3) \land c_3 \notin \{c_1, c_2\}\]

\[\exists c_4: \text{chair}(c_4) \land c_3 \notin \{c_1, c_2, c_3\}\]

\[\ldots\]
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 $r_2$
Observe a triangle and a circle touching. What is the probability the triangle is green?

\[
P(\text{green}(x) \\
| \text{triangle}(x) \land \exists y \ \text{circle}(y) \land \text{touching}(x, y))
\]

The answer depends on how the \( x \) and \( y \) were chosen!
Protocol for Observing

\[
P(\text{green}(x)) \quad | \quad \text{triangle}(x) \land \exists y \; \text{circle}(y) \land \text{touching}(x, y))
\]

<table>
<thead>
<tr>
<th>select(x)</th>
<th>select(y)</th>
<th>select(x, y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>select(y)</td>
<td>select(x)</td>
<td></td>
</tr>
</tbody>
</table>

3/4 \quad 2/3 \quad 4/5
Conclusion

- To decide what to do an agent should take into account its uncertainty and its preferences (utility).
- The field of “statistical relational AI” looks at how to combine first-order logic and probabilistic reasoning.
- We need models that can condition on observations that follow some protocol

Challenges

- **Representation**: heuristically and epistemologically adequate representations for probabilistic models + observations (+ actions + utilities + ontologies)
- **Inference**: compute posterior probabilities (or optimal actions) quickly enough to be useful
- **Learning**: get best hypotheses conditioned on all observations possible
AI: computational agents that act intelligently

What should an agent do?

**Tasks**
- Acting
- Perceiving
- Modelling
- Diagnosis
- Knowledge Acquisition
- Inference
- Learning
- Design

**Inputs**
- Ontologies
- Prior Knowledge
- Observations
- Data
- Relations
- Hypotheses
- Preferences/Utilities
- Abilities

**Foundations**
- Logic
- Probability
- Decision Theory
- Statistics
- Computation
- Game Theory
- Knowledge Representation
- Dynamical Systems

**Logic and Probability**
- Inference
- Weighted
- Existence

David Poole
Logic, Probability and Computation