Representation Issues

Luc De Raedt, Kristian Kersting,Sriraam Natarajan, David Poole

Belgium, Germany, USA, Canada

February 2017

Outline

1. Representation Issues
   - Desiderata
2. Relational models are sometimes weird
   - Directed vs undirected models
   - Population Growth
   - Varying Populations
3. What we can't do
   - Existence and Identity Uncertainty
   - Semantic Trees
   - Observation Protocols

Desiderata for a Representation

- Expressiveness:
  Is it expressive enough to solve problem at hand?

Desiderata for a Representation

- Expressiveness:
  Is it expressive enough to solve problem at hand?
- Efficient Inference:
  Is it efficient in the worst case or average case?
  Can it exploit structure (e.g., independencies and symmetries)

Desiderata for a Representation

- Expressiveness:
  Is it expressive enough to solve problem at hand?
- Efficient Inference:
  Is it efficient in the worst case or average case?
  Can it exploit structure (e.g., independencies and symmetries)
- Understandability or explainability:
  Can people understand the model?
  Can a particular prediction be explained?

Desiderata for a Representation

- Expressiveness:
  Is it expressive enough to solve problem at hand?
- Efficient Inference:
  Is it efficient in the worst case or average case?
  Can it exploit structure (e.g., independencies and symmetries)
- Understandability or explainability:
  Can people understand the model?
  Can a particular prediction be explained?
- Learnability: Can it be learned from:
  - heterogenous data
  - prior knowledge

Desiderata for a Representation

- Expressiveness:
  Is it expressive enough to solve problem at hand?
- Efficient Inference:
  Is it efficient in the worst case or average case?
  Can it exploit structure (e.g., independencies and symmetries)
- Understandability or explainability:
  Can people understand the model?
  Can a particular prediction be explained?
- Learnability: Can it be learned from:
  - heterogenous data
  - prior knowledge
- Modularity:
  Can independently developed parts be combined to form larger model?
  Can a larger model be decomposed into smaller parts?

Outline

1. Representation Issues
   - Desiderata

2. Relational models are sometimes weird
   - Directed vs undirected models
   - Population Growth
   - Varying Populations

3. What we can't do
   - Existence and Identity Uncertainty
   - Semantic Trees
   - Observation Protocols

Directed vs Undirected Models

- Undirected models (Markov networks, factor graphs) represent probability distributions in terms of factors.
  - a factor is a non-negative function of a set of variables
  - variables in a factor are neighbours of each other
  - each variable in independent of its non-neighbours given its neighbours.

Directed vs Undirected Models

- Undirected models (Markov networks, factor graphs) represent probability distributions in terms of factors.
  - a factor is a non-negative function of a set of variables
  - variables in a factor are neighbours of each other
  - each variable in independent of its non-neighbours given its neighbours.
- In directed models, factors represent conditional probabilities:
  - how each variable depends on its parents
  - each variable in independent of its non-descendents given its parents.

Directed vs Undirected Models

- Undirected models (Markov networks, factor graphs) represent probability distributions in terms of factors.
  - a factor is a non-negative function of a set of variables
  - variables in a factor are neighbours of each other
  - each variable in independent of its non-neighbours given its neighbours.
- In directed models, factors represent conditional probabilities:
  - how each variable depends on its parents
  - each variable in independent of its non-descendents given its parents.
- $\{directed\_models\} \subset \{undirected\_models\}$

Directed vs Undirected Models

- Undirected models (Markov networks, factor graphs) represent probability distributions in terms of factors.
  - a factor is a non-negative function of a set of variables
  - variables in a factor are neighbours of each other
  - each variable in independent of its non-neighbours given its neighbours.
- In directed models, factors represent conditional probabilities:
  - how each variable depends on its parents
  - each variable in independent of its non-descendents given its parents.
- $\{directed\_models\} \subset \{undirected\_models\}$
  Algorithms developed for undirected models work for both.

Directed vs Undirected Models

- Undirected models (Markov networks, factor graphs) represent probability distributions in terms of factors.
  - a factor is a non-negative function of a set of variables
  - variables in a factor are neighbours of each other
  - each variable in independent of its non-neighbours given its neighbours.
- In directed models, factors represent conditional probabilities:
  - how each variable depends on its parents
  - each variable in independent of its non-descendents given its parents.
- $\{directed\_models\} \subset \{undirected\_models\}$
  Algorithms developed for undirected models work for both.
  That does not mean that representations for undirected models can represent directed models.

Modularity

- Directed models are inherently modular.
  $P(a \mid b(X))$ is defined so that distribution over $b(c_1) \dots b(c_n)$ is not affected.

Modularity

- Directed models are inherently modular.
  $P(a \mid b(X))$ is defined so that distribution over $b(c_1) \dots b(c_n)$ is not affected.
- MLNs are provably not modular: If there is a distribution over $b(c_1) \dots b(c_n)$ (e.g., they are independent),
  $P(a \mid b(X))$ cannot be defined in an MLN so that
  - a depends on the b's $(P(a \mid b(X)) \neq P(a))$ and
  - if a is summed out, the distribution over $b(c_1) \dots b(c_n)$ is not changed.

Modularity

- Directed models are inherently modular.
  $P(a \mid b(X))$ is defined so that distribution over $b(c_1) \ldots b(c_n)$ is not affected.
- MLNs are provably not modular: If there is a distribution over $b(c_1) \ldots b(c_n)$ (e.g., they are independent),
  $P(a \mid b(X))$ cannot be defined in an MLN so that
  - a depends on the b's ($P(a \mid b(X)) \neq P(a)$) and
  - if a is summed out, the distribution over $b(c_1) \ldots b(c_n)$ is not changed.
  - Why? requires factors on arbitrary subsets of $b(x_1) \ldots b(x_k)$ — can't marry the parents

Cyclic Models

Whether people smoke depends on whether their friends smoke.

- MLN:

$$w : smokes(X) \leftarrow friends(X, Y) \wedge smokes(Y)$$

Cyclic Models

Whether people smoke depends on whether their friends smoke.

- MLN:

$$w : smokes(X) \leftarrow friends(X, Y) \wedge smokes(Y)$$

(where $\leftarrow$ is material implication) is equivalent to

$$\begin{aligned} w :& true(X) \wedge true(Y) \\ -w :& \neg smokes(X) \wedge friends(X, Y) \wedge smokes(Y) \end{aligned}$$

Cyclic Models

Whether people smoke depends on whether their friends smoke.

- MLN:

$$w : smokes(X) \leftarrow friends(X, Y) \wedge smokes(Y)$$

(where $\leftarrow$ is material implication) is equivalent to

$$\begin{aligned} w :& true(X) \wedge true(Y) \\ -w :& \neg smokes(X) \wedge friends(X, Y) \wedge smokes(Y) \end{aligned}$$

- Problog

$$w : smokes(X) \leftarrow \quad friends(X, Y) \wedge smokes(Y)$$

Cyclic Models

Whether people smoke depends on whether their friends smoke.

- MLN:

$$w : smokes(X) \leftarrow friends(X, Y) \wedge smokes(Y)$$

(where $\leftarrow$ is material implication) is equivalent to

$$\begin{aligned} w :& true(X) \wedge true(Y) \\ -w :& \neg smokes(X) \wedge friends(X, Y) \wedge smokes(Y) \end{aligned}$$

- Problog

$$w : smokes(X) \leftarrow \exists Y \; friends(X, Y) \wedge smokes(Y)$$

Cyclic Models

Whether people smoke depends on whether their friends smoke.

- MLN:

$$w : smokes(X) \leftarrow friends(X,Y) \land smokes(Y)$$

(where $\leftarrow$ is material implication) is equivalent to

$$\begin{aligned} w :& true(X) \land true(Y) \\ -w :& \neg smokes(X) \land friends(X,Y) \land smokes(Y) \end{aligned}$$

- Problog

$$w : smokes(X) \leftarrow \exists Y\ friends(X,Y) \land smokes(Y)$$

  - probability of smokes goes up as the number of friends increases!

Cyclic Models

Whether people smoke depends on whether their friends smoke.

- MLN:

$$w : smokes(X) \leftarrow friends(X, Y) \wedge smokes(Y)$$

(where $\leftarrow$ is material implication) is equivalent to

$$\begin{aligned} w :& true(X) \wedge true(Y) \\ -w :& \neg smokes(X) \wedge friends(X, Y) \wedge smokes(Y) \end{aligned}$$

- Problog

$$w : smokes(X) \leftarrow \exists Y \ friends(X, Y) \wedge smokes(Y)$$

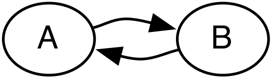
  - probability of smokes goes up as the number of friends increases!
  - Problog cannot represent negative effects: someone is less likely to smoke if their friends smoke (without there being a non-zero probability of logical inconsistency)

Cyclic Models

- Make model acyclic, by totally ordering variables.
  Destroys exchangeability. Symmetries are not preserved.

Cyclic Models

- Make model acyclic, by totally ordering variables. Destroys exchangeability. Symmetries are not preserved.
- (Relational) dependency networks: directed model,



  - $P(A, B)$ has 3 degrees of freedom,
  - $P(A \mid B), P(B \mid A)$, uses 4 numbers; typically inconsistent.
  - resulting distribution means fixed point of Markov chain.

Outline

1. Representation Issues
   - Desiderata
2. Relational models are sometimes weird
   - Directed vs undirected models
   - Population Growth
   - Varying Populations
3. What we can't do
   - Existence and Identity Uncertainty
   - Semantic Trees
   - Observation Protocols

Example

Weighted formulae:

$$-5 : \mathit{funFor}(X)$$
$$10 : \mathit{funFor}(X) \land \mathit{knows}(X, Y) \land \mathit{social}(Y)$$

If Π includes observations for all $\mathit{knows}(X, Y)$ and $\mathit{social}(Y)$:

$$P(\mathit{funFor}(X) \mid \Pi) = \mathit{sigmoid}(-5 + 10 n_T)$$

n_T is the number of individuals Y for which $\mathit{knows}(X, Y) \land \mathit{social}(Y)$ is *True* in Π.

$$\mathit{sigmoid}(x) = \frac{1}{1 + e^{-x}}$$

Example

Weighted formulae:

$$-5 : funFor(X)$$
$$10 : funFor(X) \wedge knows(X, Y) \wedge social(Y)$$

If Π includes observations for all $knows(X, Y)$ and $social(Y)$:

$$P(funFor(X) \mid \Pi) = sigmoid(-5 + 10n_T)$$

n_T is the number of individuals Y for which $knows(X, Y) \wedge social(Y)$ is *True* in Π.

$$sigmoid(x) = \frac{1}{1 + e^{-x}}$$

Using wighted formulae to define conditional probabilities is called relational logistic regression (RLR).

Abstract Example

$\alpha_0 : q$
$\alpha_1 : q \wedge \neg r(x)$
$\alpha_2 : q \wedge r(x)$
$\alpha_3 : r(x)$

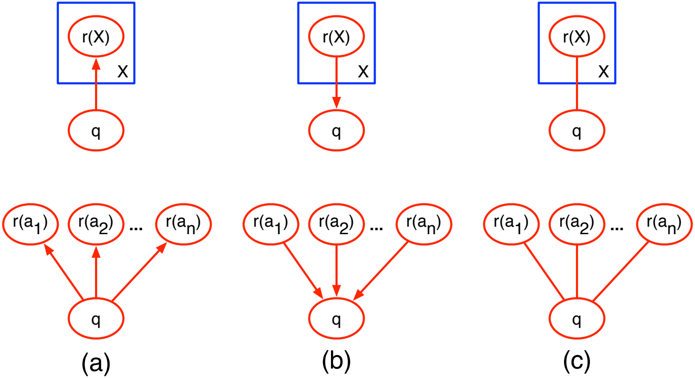
If $r(x)$ for every individual x is observed:

$$P(q \mid obs) = sigmoid(\alpha_0 + n_F\alpha_1 + n_T\alpha_2)$$

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

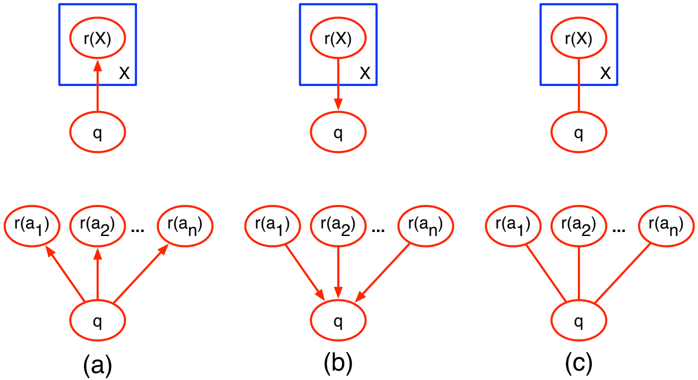
$$sigmoid(x) = \frac{1}{1 + e^{-x}}$$

Three Elementary Models



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

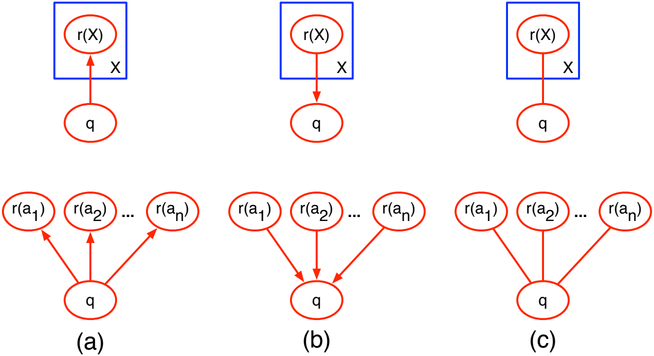
Three Elementary Models



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

— alertThey are identical models when all r's are observed.

Independence Assumptions



- Naïve Bayes (a) and Markov network (c): $R(a_i)$ and $R(a_j)$
  - are independent given Q
  - are dependent not given Q.
- Directed model with aggregation (b): $R(a_i)$ and $R(a_j)$
  - are dependent given Q,
  - are independent not given Q.

Outline

1. Representation Issues
   - Desiderata
2. Relational models are sometimes weird
   - Directed vs undirected models
   - Population Growth
   - Varying Populations
3. What we can't do
   - Existence and Identity Uncertainty
   - Semantic Trees
   - Observation Protocols

What happens as Population size n Changes: Simplest case

$$
\begin{aligned}
\alpha_0 &: q \\
\alpha_1 &: q \wedge \neg r(x) \\
\alpha_2 &: q \wedge r(x) \\
\alpha_3 &: r(x)
\end{aligned}
$$

Weighted formula define distribution:

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

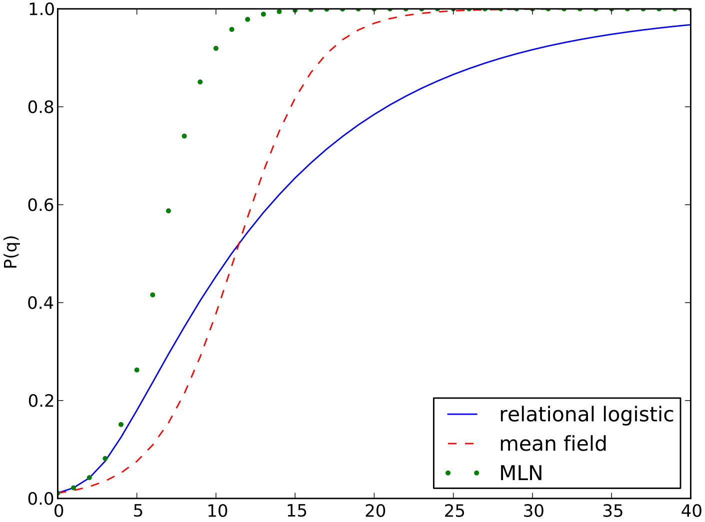
Weighted formula define conditionals:

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

Mean-field approximation:

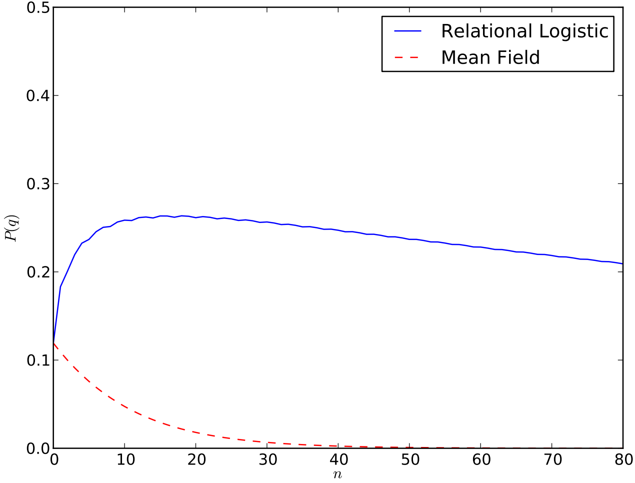
$$P_{MF}(q \mid n) = sigmoid(\alpha_0 + np_r\alpha_1 + n(1-p_r)\alpha_2)$$

Population Growth: $P(q \mid n)$

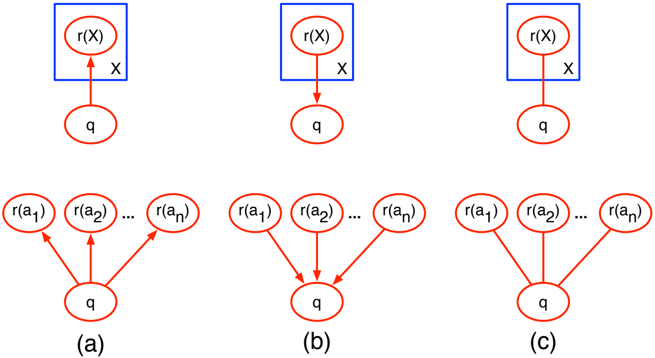


Population Growths: $P_{RLR}(q \mid n)$

Whereas this MLN is a sigmoid of n, RLR needn't be monotonic:

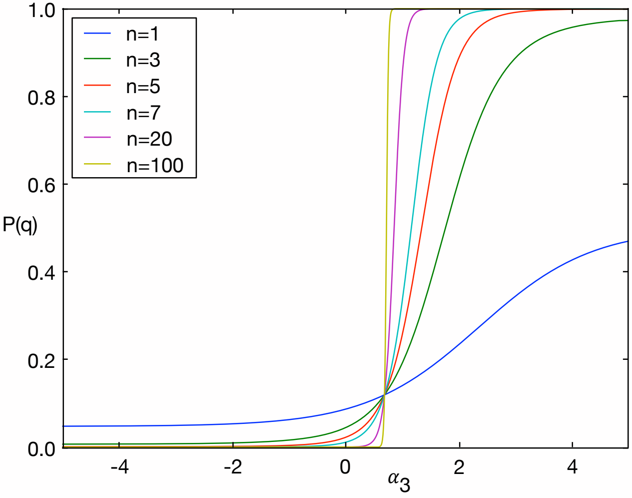


Dependence of $R(x)$ on population size

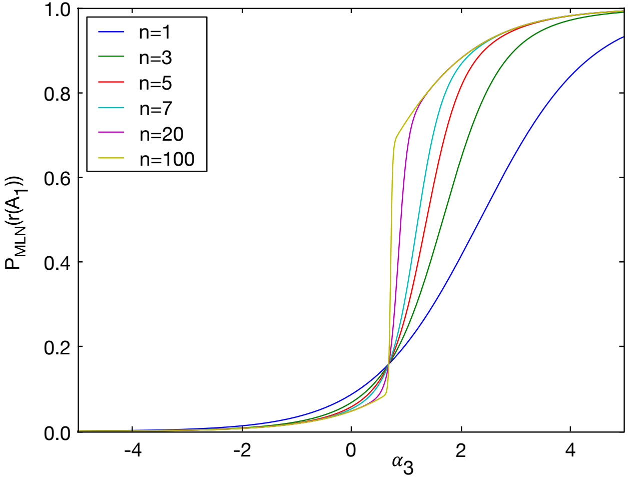


- In (b), the directed model with aggregation, $P(R(x))$ is not affected by the population size.
- In (c), $P_{MLN}(R(x))$ is unaffected by population size if and only if the MLN is equivalent to a Naïve Bayes model (a).
- For other MLNs...

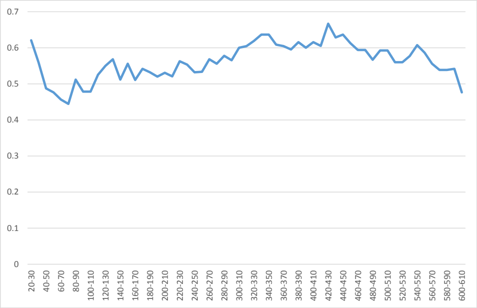
$P_{MLN}(q \mid \alpha_3)$ for various n



$P_{MLN}(r(A_1) \mid \alpha_3)$ for various n

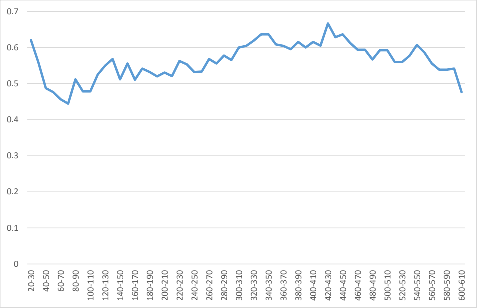


Real Data



Observed $P(25 < Age(p) < 45 \mid n)$, where n is number of movies watched from the Movielens dataset.

Real Data



Observed $P(25 < Age(p) < 45 \mid n)$, where n is number of movies watched from the Movielens dataset.
Dont use:

$$w : age(P) \leftarrow rated(P, M) \wedge foo(M)$$

then $age(P) \rightarrow \pm\infty$ as number of movies increases.

Example of polynomial dependence of population

$\alpha_0 : q$
$\alpha_1 : q \wedge true(X)$
$\alpha_2 : q \wedge r(X)$
$\alpha_3 : true(X)$
$\alpha_4 : r(X)$
$\alpha_5 : q \wedge true(X) \wedge true(Y)$
$\alpha_6 : q \wedge r(X) \wedge true(Y)$
$\alpha_7 : q \wedge r(X) \wedge r(Y)$

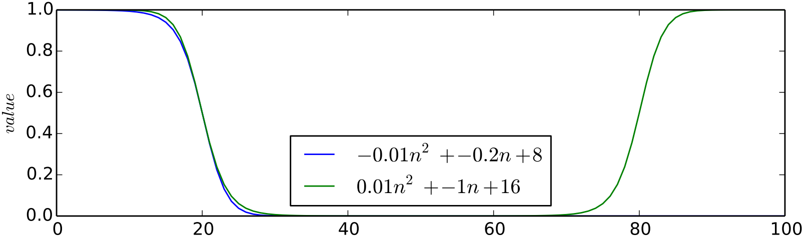
In RLR and in MLN, if all $R(A_i)$ are observed:

$$P(q \mid obs) = sigmoid(\alpha_0 + n\alpha_1 + n_T\alpha_2 + n^2\alpha_5 + n_T n\alpha_6 + n_T^2\alpha_7)$$

$R(X)$ is true for n_T individuals out of a population of n.

Danger of fitting to data without understanding the model

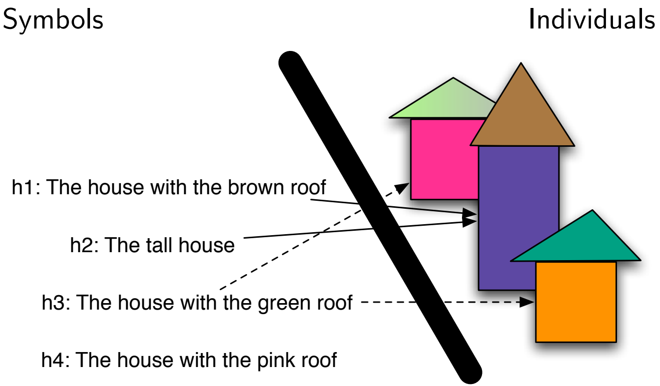
- RLR can fit sigmoid of any polynomial.
- Consider a polynomial of degree 2:



Outline

1. Representation Issues
   - Desiderata
2. Relational models are sometimes weird
   - Directed vs undirected models
   - Population Growth
   - Varying Populations
3. What we can't do
   - Existence and Identity Uncertainty
   - Semantic Trees
   - Observation Protocols

Correspondence Problem



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

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?

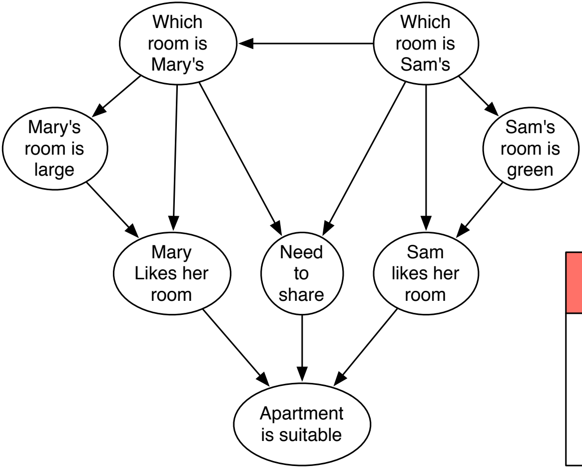
Role assignments

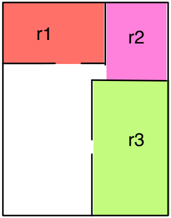
Hypothesis about what apartment Mary would like.

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

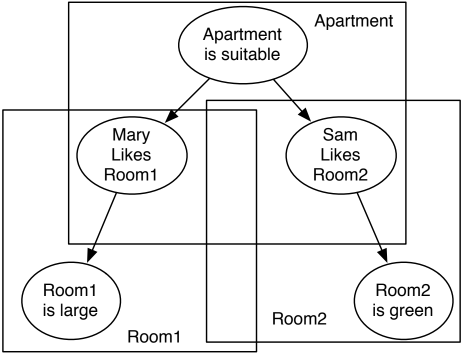
Bayesian Belief Network Representation

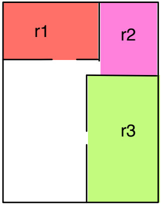




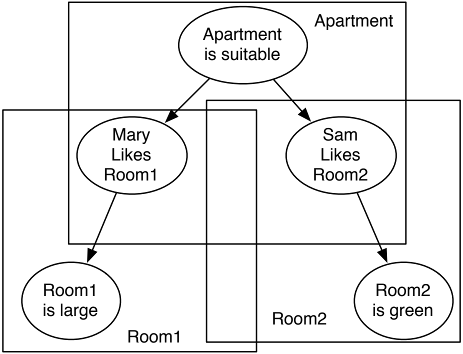
How can we condition on the observation of the apartment?

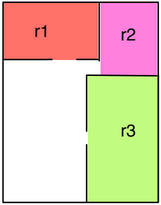
Naive Bayes representation





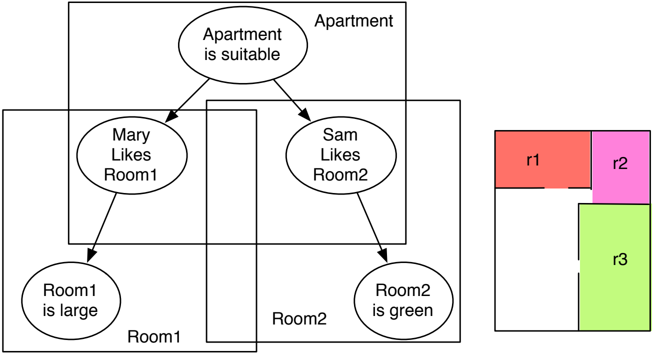
Naive Bayes representation





How do we specify that Mary chooses a room?

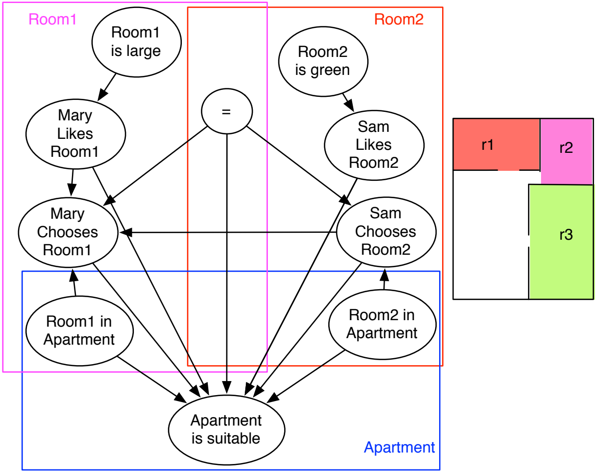
Naive Bayes representation



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

- We need more work on integrating probabilistic models with rich observations

Causal representation



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

Data

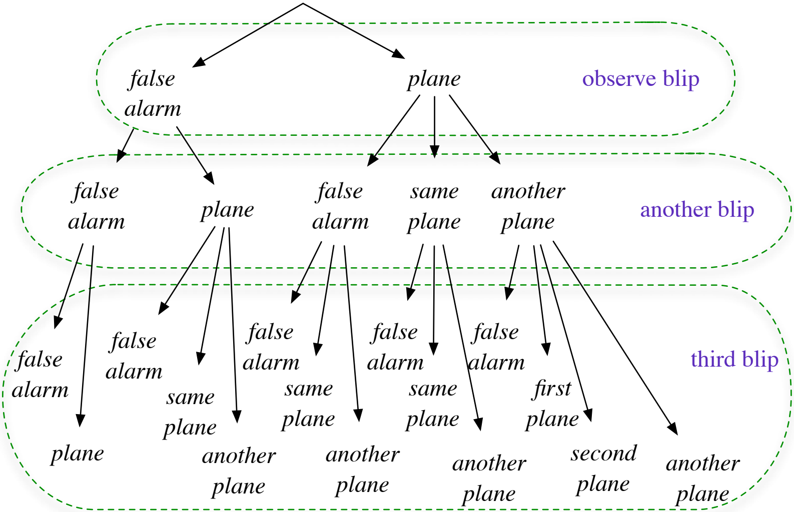
Real data is messy!

- Multiple levels of abstraction
- Multiple levels of detail
- Uses the vocabulary from many ontologies
- 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?
  - What is the provenance of the data; what was done to it when?
- Errors, forgeries, . . .

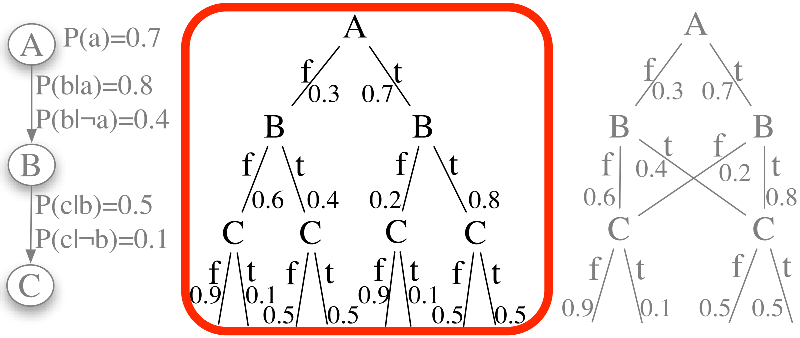
Number and Existence Uncertainty

- 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



Semantic Tree



↑

semantic tree

event tree

decision tree. . .

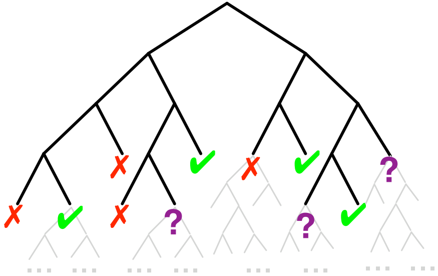
Semantic tree

- Nodes are propositions
- Left branch is when proposition is false
  Right branch is when proposition is true
- There is a probability distribution over the children of each node
- Each finite path from the root corresponds to a formula
- Each finite path from the root has a probability that is the product of the probabilities in the path

A generative model generates a semantic tree.

Infinite Semantic Tree

Given a proposition α:

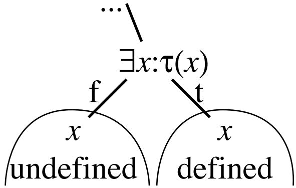


✔ *path* $\models \alpha$
✗ *path* $\models \neg\alpha$
? otherwise

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

First-order Semantic Trees

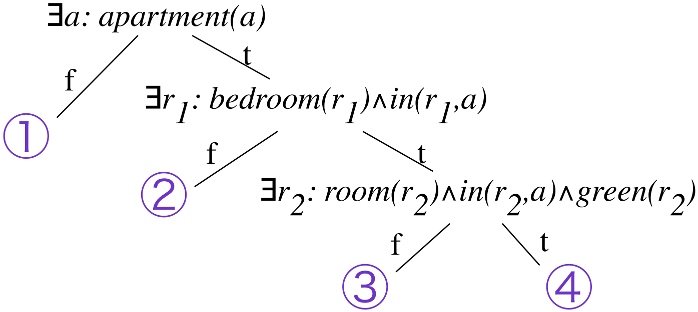
Split on quantified first-order formulae:



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

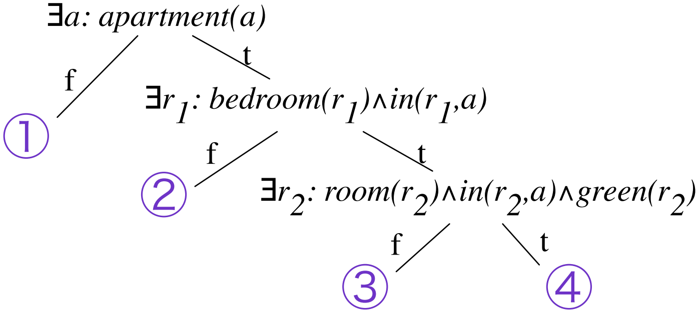
A logical generative model generates a first-order semantic tree.

First-order Semantic Tree (cont)



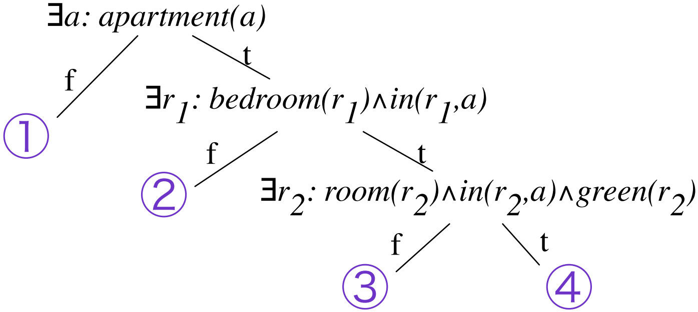
①

First-order Semantic Tree (cont)



1. there is no apartment
2.

First-order Semantic Tree (cont)

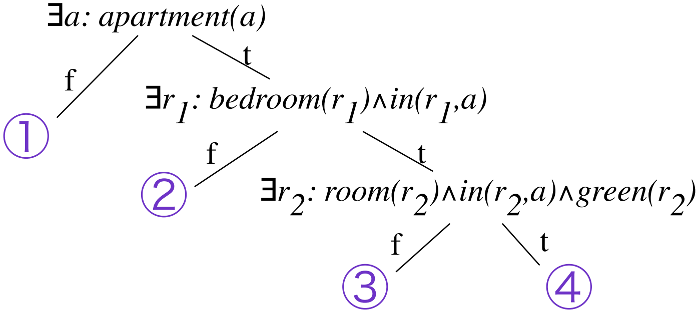


① there is no apartment

② there is no bedroom in the apartment

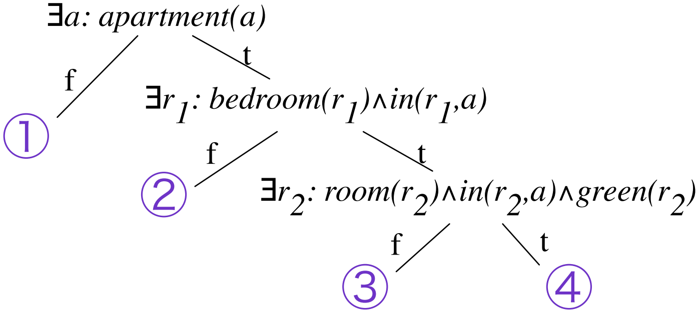
③

First-order Semantic Tree (cont)



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

First-order Semantic Tree (cont)



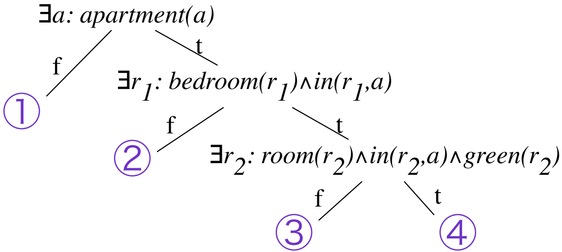
1. there is no apartment
2. there is no bedroom in the apartment
3. there is a bedroom but no green room
4. there is a bedroom and a green room

Semantics

Each path from the root corresponds to a logical formula. The **path formula** to node n is:

- The path formula of the root node is "*true*".
- If the path formula of node n is formula f and node n is labelled with formula f'
  - the "true" child of node n has path formula
    $$f \wedge f'$$
    where f' is in the scope of the quantification of f.
  - The "false" child of node n has path formula:
    $$f \wedge \neg(f \wedge f')$$

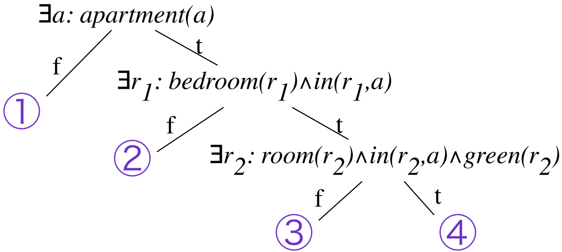
First-order Semantic Tree (cont)



Path formulae:

①

First-order Semantic Tree (cont)

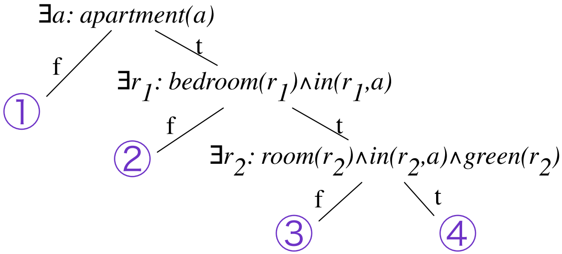


Path formulae:

① $(\neg\exists a\; apt(a))$

②

First-order Semantic Tree (cont)



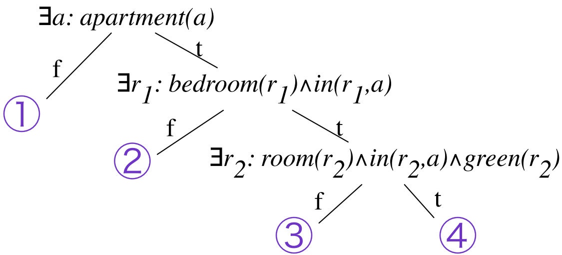
Path formulae:

① $(\neg \exists a \; apt(a))$

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

④

First-order Semantic Tree (cont)



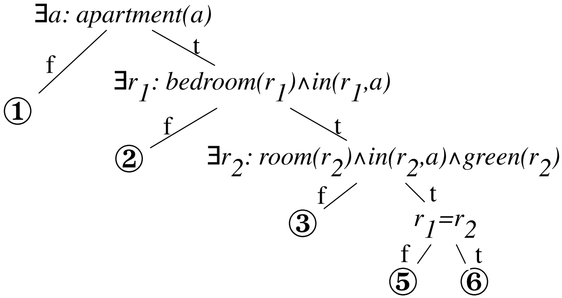
Path formulae:

① $(\neg \exists a \ apt(a))$

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

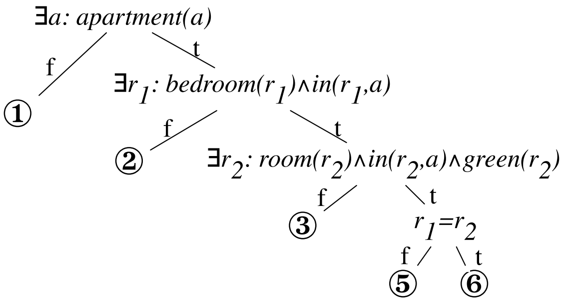
④ $\exists a \ apt(a) \wedge \exists r_1 \ br(r_1) \wedge in(r_1, a) \wedge \exists r_2 \ room(r_2) \wedge in(r_2, a) \wedge green(r_2)$

First-order Semantic Tree (cont)



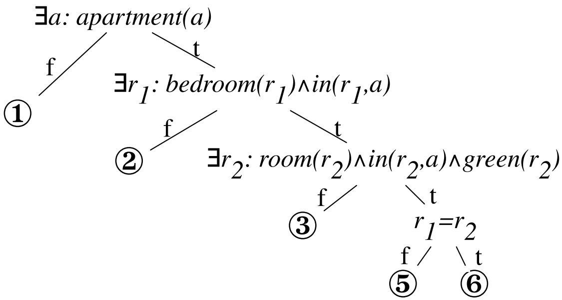
⑥

First-order Semantic Tree (cont)



⑥ $\exists a\ apt(a) \wedge \exists r_1\ br(r_1) \wedge in(r_1, a) \wedge \exists r_2\ room(r_2) \wedge in(r_2, a) \wedge green(r_2) \wedge r_1 = r_2$
means

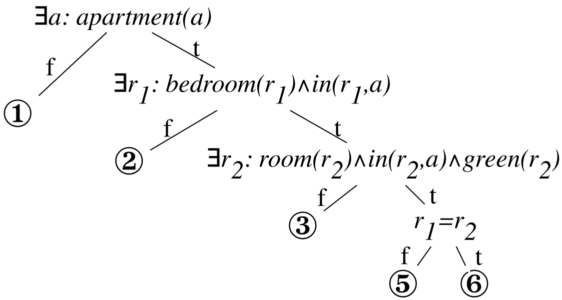
First-order Semantic Tree (cont)



⑥ $\exists a\ apt(a) \wedge \exists r_1\ br(r_1) \wedge in(r_1, a) \wedge \exists r_2\ room(r_2) \wedge in(r_2, a) \wedge green(r_2) \wedge r_1 = r_2$
means there is a green bedroom.

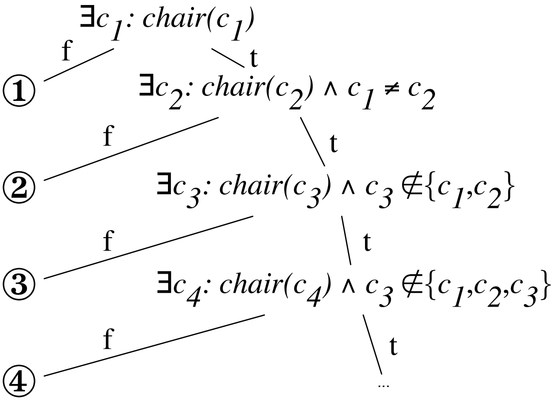
⑤

First-order Semantic Tree (cont)

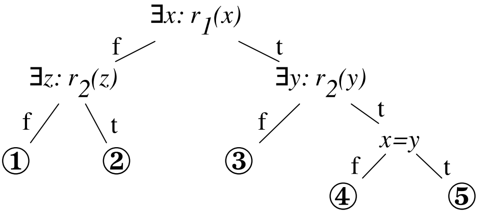


- ⑥ $\exists a\ apt(a) \wedge \exists r_1\ br(r_1) \wedge in(r_1, a) \wedge \exists r_2\ room(r_2) \wedge in(r_2, a) \wedge green(r_2) \wedge r_1 = r_2$
  means there is a green bedroom.
- ⑤ There is a bedroom and a green room, but no green bedroom.

Distributions over number

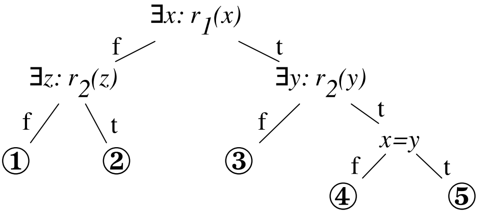


Roles and Identity (1)



①

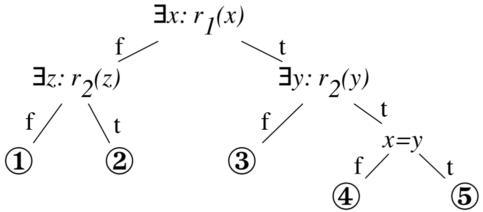
Roles and Identity (1)



① there no individual filling either role

②

Roles and Identity (1)

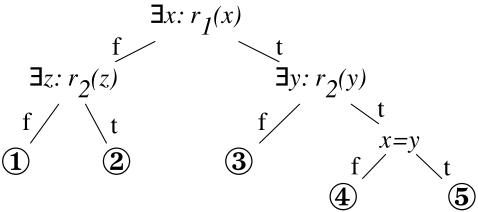


① there no individual filling either role

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

③

Roles and Identity (1)



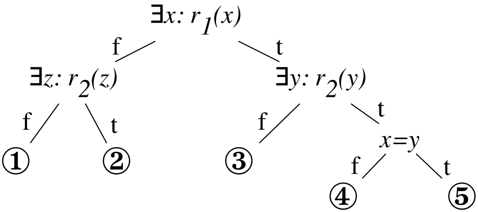
① there no individual filling either role

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

Roles and Identity (1)



① there no individual filling either role

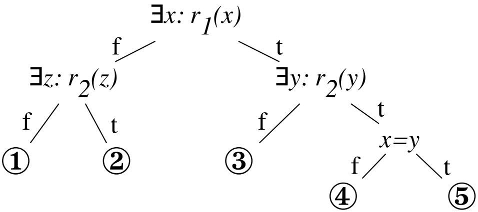
② there is an individual filling role r_2 but none filling r_1

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

Roles and Identity (1)



① there no individual filling either role

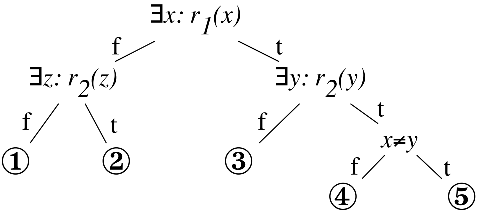
② there is an individual filling role r_2 but none filling r_1

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

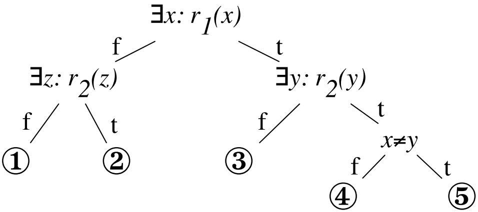
⑤ some individual fills both roles r_1 and r_2

Roles and Identity (2)



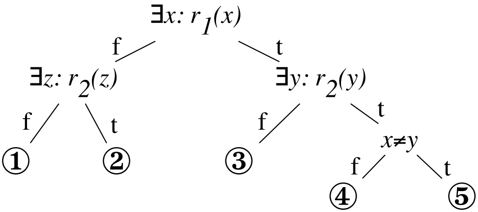
①

Roles and Identity (2)



1. there no individual filling either role
2.

Roles and Identity (2)

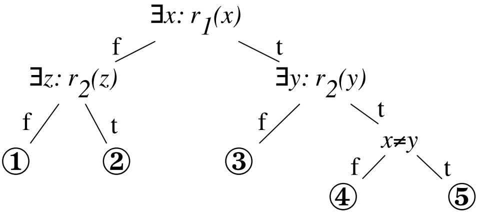


① there no individual filling either role

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

③

Roles and Identity (2)



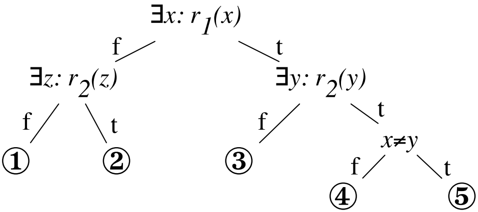
① there no individual filling either role

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

Roles and Identity (2)



① there no individual filling either role

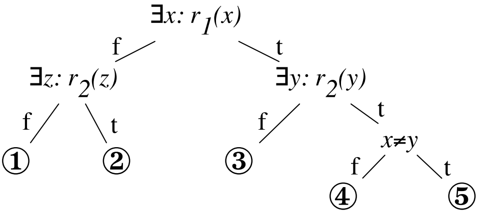
② there is an individual filling role r_2 but none filling r_1

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

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

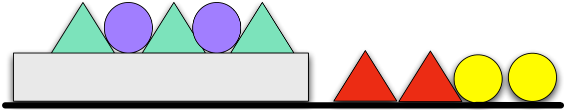
④ only the same individual fill roles r_1 and r_2

⑤ there are different individuals that fill roles r_1 and r_2

Outline

1. Representation Issues
   - Desiderata
2. Relational models are sometimes weird
   - Directed vs undirected models
   - Population Growth
   - Varying Populations
3. What we can't do
   - Existence and Identity Uncertainty
   - Semantic Trees
   - Observation Protocols

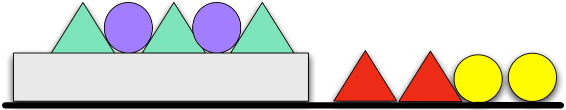
Observation Protocols



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

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

Observation Protocols



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

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

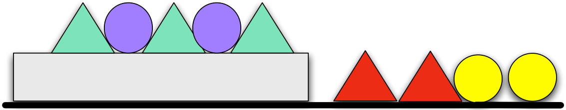
The answer depends on how the x and y were chosen!

Exchangeability

- Exchangeability: a priori each individual is equally likely to be chosen.
- A generalized first-order semantic tree is a first-order semantic tree that can contain $commit(\overline{x})$ nodes.
  For each $commit(\overline{x})$ node:
  - $\overline{x}$ is a set of variables
  - the node is in the scope of each x in $\overline{x}$
  - no x is in an ancestor commit.
  - this node has one child.

  For each possible world, each tuple of individuals that satisfies the path formula to $commit(\overline{x})$ has an equal chance of being chosen.

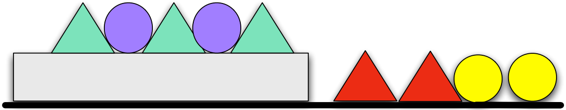
Protocol for Observing



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

$select(x)$	$select(y)$	$select(x,y)$
$select(y)$	$select(x)$	

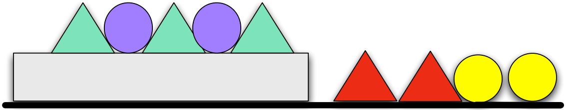
Protocol for Observing



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

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

Protocol for Observing



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

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

A logical formula does not provide enough information to determine the probabilities.

Challenges

- Heterogeneity: information about individuals varies greatly in kind and amount (e.g., information in patients' electronic health records, number of movies people have rated)
- Representations should
  - let people state their prior knowledge,
  - let them understand what they stated, and
  - let them understand the posterior models (given evidence).
- Use the meta-data of how data was collected
- Models often refer to roles that are not observed