Logic programming

- Prolog, Alain Colmerauer 1972
- Program = a knowledge base expressed as definite clauses
- Queries to the knowledge base
- **Closed world assumption:** we assume \( \neg \varphi \) to be true if sentence \( \varphi \) is not entailed by the knowledge base
- Syntax:
  - Capital characters denote variables,
  - Small character stand for constants,
  - The head of the rule precedes the body,
  - Instead of implication use \( :\rightarrow \),
  - Comma stand for conjunction,
  - Period ends a sentence

\[
\text{thesis}_2010(X) :\rightarrow \text{student}(X), \text{eager\_to\_learn}(X).
\]
- Prolog has a lot of syntactic sugar, e.g., for lists and arithmetics

Prolog program

\[
\text{append}(X,Y,Z) \text{ succeeds if list } Z \text{ is the result of appending (catenating) lists } X \text{ and } Y
\]

\[
\begin{align*}
\text{append}([],Y,Y). \\
\text{append}([A|X],Y,[A|Z]):\rightarrow \text{append}(X,Y,Z).
\end{align*}
\]

Query:

\[
\text{append}([1],[2],Z)?
\]

\( Z=[1,2] \)

- We can also ask the query

\[
\text{append}(A,B,[1,2])?:
\]

Appending what two lists gives the list \([1,2]\)?

- As the answer we get back all possible substitutions

\[
\begin{align*}
A=[ ] &\quad B=[1,2] \\
A=[1,2] &\quad B=[ ]
\end{align*}
\]
The clauses in a Prolog program are tried in the order in which they are written in the knowledge base. Also the conjuncts in the body of the clause are examined in order (from left to right). There is a set of built-in functions for arithmetic, which need not be inferred further. E.g., `X is 4+3 → X=7`

For instance I/O is taken care of using built-in predicates that have side effect when executed.

Negation as failure

```
alive(X) :- not dead(X).

"Everybody is alive if not provably dead"
```

The negation in Prolog does not correspond to the negation of logic (using the closed world assumption)

```
single_student(X) :-
    not married(X), student(X).
student(peter).
married(john).
```

By the closed world assumption, `X=peter` is a solution to the program.

The execution of the program, however, fails because when `X=john` the first predicate of the body fails.

If the conjuncts in the body were inverted, it would succeed.
• An equality goal succeeds if the two terms are unifiable
  • E.g., \( X+Y = 2+3 \rightarrow X=2, \ Y=3 \)

• Prolog omits some necessary checks in connection of variable bindings \( \rightarrow \) Inference is not sound
• These are seldom a problem
• Depth-first search can lead to infinite loops (= incomplete)
  \[
  \text{path}(X,Z) :- \text{path}(X,X), \text{link}(Y,Z).
  \]
  \[
  \text{path}(X,Z) :- \text{link}(X,Z).
  \]

• Careful programming, however, lets us escape such problems
  \[
  \text{path}(X,Z) :- \text{link}(X,Z).
  \]
  \[
  \text{path}(X,Z) :- \text{path}(X,Y), \text{link}(Y,Z).
  \]

• Anonymous variable
  \[
  \text{member}(X,[X|\_]).
  \]
  \[
  \text{member}(X,[\_|Y]) :- \text{member}(X,Y).
  \]

• Works just fine
  \[
  \text{member}(d,[a,b,c,d,e,f,g])?
  \]
  yes
  \[
  \text{member}(2,[3,a,4,f])?
  \]
  no

• But queries
  \[
  \text{member}(a,X)\
  \text{member}(a,[a,b,r,a,c,a,d,a,b,r,a])?
  \]
  do not necessarily give the intended answers
• We can explicitly prune the execution of Prolog programs by cutting.
• Negation using cut
  
  not X:- X, !, fail.
  not X.

• fail causes the program to fail
• At the point of a cut all bindings that have been made since starting to examine the rule are fixed
• For the conjuncts in the body preceding the cut, no new solutions are searched for
• Neither does one examine other rules having the same head.

• Prolog may come up with the same answer through several inference paths
• Then the same answer is returned more than once
  
  minimum(X,Y,X):- X<=Y.
  minimum(X,Y,Y):- X>=Y.

• Both rules yield the same answer for the query minimum(2,2,M) ?
• One must be careful in using cut for optimizing inference
  
  minimum(X,Y,X):- X<=Y, !.
  minimum(X,Y,Y).

• This program is erroneous, for instance minimum(2,8,8) holds according to it.
• The key question of Prolog and logic programming obviously is efficiency of execution
• Prolog implementations use a wide variety of enhancement techniques
  • For example, instead of generating all possible solutions for a subgoal before examining the next subgoal, a Prolog interpreter is content (so far) with just one
  • Similarly variable binding is at each instant unique; only when the search runs into a dead end, can backing up to a choice point lead to unbinding of variables
• A stack of history, called the trail, needs to be maintained to keep track of all variable bindings

9.5 Resolution

• Kurt Gödel's completeness theorem (1930) for first-order logic: any entailed sentence has a finite proof
  \( T 
  \equiv
  \phi
  \iff
  T
  
  \equiv
  \phi
  \)

• It was not until Robinson's (1965) resolution algorithm that a practical proof procedure was found
• Gödel's more famous result is the incompleteness theorem: a logical system that includes the principle of induction, is necessary incomplete
• There are sentences that are entailed, but have no finite proof
• This holds in particular for number theory, which thus cannot be axiomatized
• For resolution, we need to convert the sentences to CNF
• E.g., “Everyone who loves all animals is loved by someone”

\[ \forall x: [\forall y: \text{Animal}(y) \iff \text{Loves}(x, y)] \iff [\exists y: \text{Loves}(y, x)]. \]

• Eliminate implications
\[ \forall x: [\neg \forall y: \neg \text{Animal}(y) \lor \text{Loves}(x, y)] \lor [\exists y: \text{Loves}(y, x)]. \]

• Move negation inwards
\[ \forall x: [\exists y: \text{Animal}(y) \land \neg \text{Loves}(x, y)] \lor [\exists y: \text{Loves}(y, x)]. \]

• Standardize variables
\[ \forall x: [\exists y: \text{Animal}(y) \land \neg \text{Loves}(x, y)] \lor [\exists z: \text{Loves}(z, x)]. \]

• Skolemization
\[ \forall x: [\text{Animal}(F(x)) \land \neg \text{Loves}(x, F(x))] \lor \text{Loves}(G(x), x). \]

• Drop universal quantifiers
\[ [\text{Animal}(F(x)) \land \neg \text{Loves}(x, F(x))] \lor \text{Loves}(G(x), x). \]

• Distribute \( \land \) over \( \lor \)
\[ [\text{Animal}(F(x)) \lor \text{Loves}(G(x), x)] \land [\neg \text{Loves}(x, F(x)) \lor \text{Loves}(G(x), x)]. \]

• The end result is quite hard to comprehend, but it doesn’t matter, because the translation procedure is easily automated
First-order literals are complementary if one unifies with the negation of the other. Thus the binary resolution rule is

\[
\frac{\ell_1 \lor \ldots \lor \ell_k \lor m}{(\ell_1 \lor \ldots \lor \ell_{i-1} \lor \ell_{i+1} \lor \ldots \lor \ell_k)(\theta)}
\]

where \(\text{Unify}(\ell_i, \neg m) = \theta\). Resolution is a complete inference rule also for predicate logic in the sense that we can check (not generate) all logical consequences of the knowledge base. KB ⊨ α is proved by showing that KB ∧ ¬α is unsatisfiable through a proof by refutation.

Theorem provers (automated reasoners) accept full first-order logic, whereas most logic programming languages handle only Horn clauses. For example Prolog intertwines logic and control. In most theorem provers, the syntactic form chosen for the sentences does not affect the result. Application areas: verification of software and hardware. In mathematics theorem provers have a high standing nowadays: they have come up with novel mathematical results. For instance, in 1996 a version of well-known Otter was the first to prove (eight days of computation) that the axioms proposed by Herbert Robbins in 1933 really define Boolean algebra.
13 UNCERTAINTY

- In practice agents almost never have full access to the whole truth of their environment and, therefore, must act under uncertainty.
- A logical agent may fail to acquire certain knowledge that it would require.
- If the agent cannot conclude that any particular course of action achieves its goal, then it will be unable to act.
- Conditional planning can overcome uncertainty to some extent, but it does not resolve it.
- An agent based solely on logics cannot choose rational actions in an uncertain environment.

- Logical knowledge representation requires rules without exceptions.
- In practice, we can typically at best provide some degree of belief for a proposition.
- In dealing with degrees of belief we will use probability theory.
- Probability 0 corresponds to an unequivocal belief that the sentence is false and, respectively, 1 to an unequivocal belief that the sentence is true.
- Probabilities in between correspond to intermediate degrees of belief in the truth of the sentence, not on its relative truth.
- Utilities that have been weighted with probabilities give the agent a chance of acting rationally by preferring the action that yields the highest expected utility.
- Principle of Maximum Expected Utility (MEU).
13.2 Basic Probability Notation

- A random variable refers to a part of the world, whose status is initially unknown.
- Random variables play a role similar to proposition symbols in propositional logic.
- E.g., Cavity might refer whether the lower left wisdom tooth has a cavity.
- The domain of a random variable may be of type:
  - Boolean: we write Cavity = true ⇔ cavity and Cavity = false ⇔ ¬cavity;
  - discrete: e.g., Weather might have the domain { sunny, rainy, cloudy, snow };
  - continuous: then one usually examines the cumulative distribution function; e.g., X ≤ 4.02

Elementary propositions can be combined to form complex propositions using all the standard connectives:

\[ \text{cavity} \land \neg \text{toothache} \]

- An atomic event is a complete specification of the world, i.e., an assignment of values to all the variables.
- Properties of atomic events:
  - They are mutually exclusive.
  - The set of all atomic events is exhaustive — at least one must be the case.
  - Any particular atomic event entails the truth or falsehood of every proposition.
  - Any proposition is logically equivalent to the disjunction of all atomic events that entail the truth of the proposition.

\[ \text{cavity} \equiv (\text{cavity} \land \text{toothache}) \lor (\text{cavity} \land \neg \text{toothache}) \]
Prior probability

- Degrees of belief are always applied to propositions
- Prior probability \( P(a) \) is the degree of belief accorded to proposition \( a \) in the absence of any other information
  \[ P(\text{cavity}) = 0.1 \]
- In particular, prior probability is the agent’s initial belief before it receives any percepts

- The probability distribution \( P(X) \) of a random variable \( X \) is a vector of values for probabilities of the elements in its (ordered) domain
  - E.g., when \( P(\text{sunny}) = 0.02, P(\text{rainy}) = 0.2, P(\text{cloudy}) = 0.7, \) and \( P(\text{snow}) = 0.08 \), then
    \[ P(\text{Weather}) = [0.02, 0.2, 0.7, 0.08] \]

• The joint probability distribution of two random variables is the product of their domains
  - E.g., \( P(\text{Weather}, \text{Cavity}) \) is a \( 4 \times 2 \) table of probabilities
  - Full joint probability distribution covers the complete set of random variables used to describe the world

- For continuous variables it is not possible to write out the entire distribution as a table, one has to examine probability density functions instead
- Rather than examine point probabilities (that have value 0), we examine probabilities of value ranges
- We will concentrate mostly on discrete-valued random variables
Conditional probability

- Once the agent has obtained some evidence concerning the previously unknown random variables making up the domain, we have to switch to using conditional (posterior) probabilities.
- \( P(a \mid b) \) is the probability of proposition \( a \), given that all we know is \( b \):
  \[
P(cavity \mid toothache) = 0.8
\]
- \( P(cavity) = P(cavity \mid ) \)
- We can express conditional probabilities in terms of unconditional probabilities:
  \[
P(a \mid b) = \frac{P(a \land b)}{P(b)}
\]
  whenever \( P(b) > 0 \)

- The previous equation can also be written as the product rule:
  \[
P(a \land b) = P(a \mid b) \cdot P(b)
\]
  
- We can, of course, have the rule the other way around
  \[
P(a \land b) = P(b \mid a) \cdot P(a)
\]
  
- Conditional distributions:
  \[
P(X \mid Y) \equiv P(X = x_i \mid Y = y_j) \quad \forall \ i, j
\]
  
- By the product rule
  \[
P(X,Y) = P(X \mid Y) \cdot P(Y)
\]
  (entry-by-entry, not a matrix multiplication)
13.3 The Kolmogorov Axioms of Probability

- The axiomatization of probability theory (1933) based on three simple axioms

1. For any proposition \( a \) the probability is in between 0 and 1:
   \[ 0 \leq P(a) \leq 1 \]

2. Necessarily true (i.e., valid) propositions have probability 1 and necessarily false (i.e., unsatisfiable) propositions have probability 0:
   \[ P(\text{true}) = 1 \quad P(\text{false}) = 0 \]

3. The probability of a disjunction is given by
   \[ P(a \lor b) = P(a) + P(b) - P(a \land b) \]

We can derive a variety of useful facts from the basic axioms; e.g.:

- \[ P(a \lor \neg a) = P(a) + P(\neg a) - P(a \land \neg a) \]
- \[ P(\text{true}) = P(a) + P(\neg a) - P(\text{false}) \]
  \[ 1 = P(a) + P(\neg a) \]
  \[ P(\neg a) = 1 - P(a) \]

- The fact of the third line can be extended for a discrete variable \( D \) with the domain \( d_1, \ldots, d_n \): 
  \[ \sum_{i=1}^{n} P(D = d_i) = 1 \]

- For a continuous variable \( X \) the summation is replaced by an integral:
  \[ \int_{-\infty}^{\infty} P(X = x) \, dx = 1 \]
• The probability distribution on a single variable must sum to 1
• It is also true that any joint probability distribution on any set of variables must sum to 1
• Recall that any proposition \(a\) is equivalent to the disjunction of all the atomic events in which \(a\) holds
• Call this set of events \(e(a)\)
• Atomic events are mutually exclusive, so the probability of any conjunction of atomic events is zero, by axiom 2
• Hence, from axiom 3
\[
P(a) = \sum_{e \in e(a)} P(e)
\]

• Given a full joint distribution that specifies the probabilities of all atomic events, this equation provides a simple method for computing the probability of any proposition

13.4 Inference Using Full Joint Distribution

<table>
<thead>
<tr>
<th></th>
<th>toothache</th>
<th>~toothache</th>
</tr>
</thead>
<tbody>
<tr>
<td>catch</td>
<td>0.108</td>
<td>0.012</td>
</tr>
<tr>
<td>~catch</td>
<td>0.072</td>
<td>0.008</td>
</tr>
<tr>
<td>cavity</td>
<td>0.016</td>
<td>0.064</td>
</tr>
<tr>
<td>~cavity</td>
<td>0.144</td>
<td>0.576</td>
</tr>
</tbody>
</table>

• E.g., there are six atomic events for \(cavity \lor toothache\): 0.108 + 0.012 + 0.072 + 0.008 + 0.016 + 0.064 = 0.28
• Extracting the distribution over a variable (or some subset of variables), marginal probability, is attained by adding the entries in the corresponding rows or columns
• E.g., \(P(cavity) = 0.108 + 0.012 + 0.072 + 0.008 = 0.2\)
• We can write the following general marginalization rule for any sets of variables \(Y\) and \(Z\):
\[
P(Y) = \sum_{z \in Z} P(Y, z)
\]
Computing a conditional probability

\[ P(\text{cavity} | \text{toothache}) = \frac{P(\text{cavity} \land \text{toothache})}{P(\text{toothache})} = \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.6 \]

Respectively

\[ P(\neg\text{cavity} | \text{toothache}) = \frac{0.08}{0.2} = 0.4 \]

\[ \frac{1}{P(\text{toothache})} = \frac{1}{0.2} = 5 \]

is a normalization constant ensuring that the distribution \( P(\text{Cavity} | \text{toothache}) \) adds up to 1

Let \( \alpha \) denote the normalization constant

\[ P(\text{Cavity} | \text{toothache}) = \alpha \frac{P(\text{Cavity, toothache})}{P(\text{toothache})} = \alpha \left( P(\text{Cavity, toothache, catch}) + P(\text{Cavity, toothache, }\neg\text{catch}) \right) = \alpha \left( [0.108, 0.016] + [0.012, 0.064] \right) = \alpha \left( [0.12, 0.08] \right) = [0.6, 0.4] \]