Logical agents

CHAPTER 6

Outline

◊ Knowledge bases
◊ Wumpus world
◊ Logic in general
◊ Propositional (Boolean) logic
◊ Normal forms
◊ Inference rules
Knowledge bases

<table>
<thead>
<tr>
<th>Inference engine</th>
<th>domain-independent algorithms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge base</td>
<td>domain-specific content</td>
</tr>
</tbody>
</table>

Knowledge base = set of sentences in a formal language

Declarative approach to building an agent (or other system):

TELL it what it needs to know

Then it can ASK itself what to do—answers should follow from the KB

Agents can be viewed at the knowledge level
i.e., what they know, regardless of how implemented

Or at the implementation level
i.e., data structures in KB and algorithms that manipulate them

A simple knowledge-based agent

```plaintext
function KB-Agent( percept ) returns an action
    static KB, a knowledge base
    t, a counter, initially 0, indicating time
    TELL( KB, MAKE-PERCEPT-SENTENCE( percept, t ) )
    action ← ASK( KB, MAKE-ACTION-QUERY( ) )
    TELL( KB, MAKE-ACTION-SENTENCE( action, t ) )
    t ← t + 1
    return action
```

The agent must be able to:
Represent states, actions, etc.
Incorporate new percepts
Update internal representations of the world
Deduce hidden properties of the world
Deduce appropriate actions
Logic in general

Logics are formal languages for representing information such that conclusions can be drawn.

Syntax defines the sentences in the language.

Semantics define the "meaning" of sentences; i.e., define truth of a sentence in a world.

E.g., the language of arithmetic:
- $x + 2 \geq y$ is a sentence; $x2 + y >$ is not a sentence.
- $x + 2 \geq y$ is true iff the number $x + 2$ is no less than the number $y$.
- $x + 2 \geq y$ is true in a world where $x = 7$, $y = 1$.
- $x + 2 \geq y$ is false in a world where $x = 0$, $y = 6$.

Types of logic

Logics are characterized by what they commit to as "primitives".


Epistemological commitment: what states of knowledge?

<table>
<thead>
<tr>
<th>Language</th>
<th>Ontological Commitment</th>
<th>Epistemological Commitment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Propositional logic</td>
<td>facts</td>
<td>true/false/unknown</td>
</tr>
<tr>
<td>First-order logic</td>
<td>facts, objects, relations</td>
<td>true/false/unknown</td>
</tr>
<tr>
<td>Temporal logic</td>
<td>facts, objects, relations, times</td>
<td>true/false/unknown</td>
</tr>
<tr>
<td>Probability theory</td>
<td>facts</td>
<td>degree of belief 0...1</td>
</tr>
<tr>
<td>Fuzzy logic</td>
<td>degree of truth</td>
<td>degree of belief 0...1</td>
</tr>
</tbody>
</table>
Entailment

$KB \models \alpha$

Knowledge base $KB$ entails sentence $\alpha$
if and only if
$\alpha$ is true in all worlds where $KB$ is true

E.g., the KB containing "the Giants won" and "the Reds won" entails "Either the Giants won or the Reds won"

Inference

$KB \vdash_i \alpha = \text{sentence } \alpha \text{ can be derived from } KB \text{ by procedure } i$

Soundness: $i$ is sound if
whenever $KB \vdash_i \alpha$, it is also true that $KB \models \alpha$

Completeness: $i$ is complete if
whenever $KB \models \alpha$, it is also true that $KB \vdash_i \alpha$

Preview: we will define a logic (first-order logic) which is expressive enough to say almost anything of interest, and for which there exists a sound and complete inference procedure.

That is, the procedure will answer any question whose answer follows from what is known by the $KB$. 
6.4 Propositional Logic: Syntax and Semantics

6.4 Propositional Logic: Models, Rules of Inference and Complexity
Propositional inference: Enumeration method

Let $\alpha = A \lor B$ and $KB = (A \lor C) \land (B \lor \neg C)$

Is it the case that $KB \models \alpha$?
Check all possible models—$\alpha$ must be true wherever $KB$ is true

<table>
<thead>
<tr>
<th>$A$</th>
<th>$B$</th>
<th>$C$</th>
<th>$A \lor C$</th>
<th>$B \lor \neg C$</th>
<th>$KB$</th>
<th>$\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>False</td>
<td>False</td>
<td>False</td>
<td>False</td>
<td>False</td>
<td>False</td>
<td>False</td>
</tr>
<tr>
<td>False</td>
<td>False</td>
<td>True</td>
<td>False</td>
<td>False</td>
<td>False</td>
<td>False</td>
</tr>
<tr>
<td>False</td>
<td>True</td>
<td>False</td>
<td>False</td>
<td>True</td>
<td>False</td>
<td>False</td>
</tr>
<tr>
<td>False</td>
<td>True</td>
<td>True</td>
<td>False</td>
<td>True</td>
<td>False</td>
<td>False</td>
</tr>
<tr>
<td>True</td>
<td>False</td>
<td>False</td>
<td>True</td>
<td>False</td>
<td>True</td>
<td>True</td>
</tr>
<tr>
<td>True</td>
<td>False</td>
<td>True</td>
<td>True</td>
<td>True</td>
<td>True</td>
<td>True</td>
</tr>
<tr>
<td>True</td>
<td>True</td>
<td>False</td>
<td>True</td>
<td>True</td>
<td>True</td>
<td>True</td>
</tr>
<tr>
<td>True</td>
<td>True</td>
<td>True</td>
<td>True</td>
<td>True</td>
<td>True</td>
<td>True</td>
</tr>
</tbody>
</table>

Propositional inference: Solution

<table>
<thead>
<tr>
<th>$A$</th>
<th>$B$</th>
<th>$C$</th>
<th>$A \lor C$</th>
<th>$B \lor \neg C$</th>
<th>$KB$</th>
<th>$\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>False</td>
<td>False</td>
<td>False</td>
<td>False</td>
<td>True</td>
<td>False</td>
<td>False</td>
</tr>
<tr>
<td>False</td>
<td>False</td>
<td>True</td>
<td>True</td>
<td>False</td>
<td>True</td>
<td>True</td>
</tr>
<tr>
<td>False</td>
<td>True</td>
<td>False</td>
<td>True</td>
<td>True</td>
<td>True</td>
<td>True</td>
</tr>
<tr>
<td>False</td>
<td>True</td>
<td>True</td>
<td>True</td>
<td>True</td>
<td>True</td>
<td>True</td>
</tr>
<tr>
<td>True</td>
<td>False</td>
<td>False</td>
<td>True</td>
<td>True</td>
<td>True</td>
<td>True</td>
</tr>
<tr>
<td>True</td>
<td>False</td>
<td>True</td>
<td>True</td>
<td>True</td>
<td>True</td>
<td>True</td>
</tr>
<tr>
<td>True</td>
<td>True</td>
<td>False</td>
<td>True</td>
<td>True</td>
<td>True</td>
<td>True</td>
</tr>
<tr>
<td>True</td>
<td>True</td>
<td>True</td>
<td>True</td>
<td>True</td>
<td>True</td>
<td>True</td>
</tr>
</tbody>
</table>
Validity and Satisfiability

A sentence is valid if it is true in all models
\[ \text{e.g., } A \lor \neg A, \quad A \Rightarrow A, \quad (A \land (A \Rightarrow B)) \Rightarrow B \]

Validity is connected to inference via the Deduction Theorem:
\[ KB \models \alpha \text{ if and only if } (KB \Rightarrow \alpha) \text{ is valid} \]

A sentence is satisfiable if it is true in some model
\[ \text{e.g., } A \lor B, \quad C \]

A sentence is unsatisfiable if it is true in no models
\[ \text{e.g., } A \land \neg A \]

Satisfiability is connected to inference via the following:
\[ KB \models \alpha \text{ if and only if } (KB \land \neg \alpha) \text{ is unsatisfiable} \]

i.e., prove \( \alpha \) by reductio ad absurdum

---

Proof methods

Proof methods divide into (roughly) two kinds:

Model checking
- truth table enumeration (sound and complete for propositional)
- heuristic search in model space (sound but incomplete)
  \[ \text{e.g., the GSAT algorithm (Ex. 6.15)} \]

Application of inference rules
- Legitimate (sound) generation of new sentences from old
  \[ \text{Proof = a sequence of inference rule applications} \]
- Can use inference rules as operators in a standard search alg.
**Inference rules for propositional logic**

Resolution (for CNF): complete for propositional logic

\[
\begin{align*}
\alpha \lor \beta, & \quad \neg\beta \lor \gamma \\
\hline
\alpha \lor \gamma
\end{align*}
\]

Modus Ponens (for Horn Form): complete for Horn KBs

\[
\begin{align*}
\alpha_1, \ldots, \alpha_n, & \quad \alpha_1 \land \cdots \land \alpha_n \Rightarrow \beta \\
\hline
\beta
\end{align*}
\]

Can be used with forward chaining or backward chaining

---

**Wumpus World PAGE description**

**Percepts** Breeze, Glitter, Smell

**Actions** Left turn, Right turn, Forward, Grab, Release, Shoot

**Goals** Get gold back to start without entering pit or wumpus square

**Environment**
- Squares adjacent to wumpus are smelly
- Squares adjacent to pit are breezy
- Glitter if and only if gold is in the same square
- Shooting kills the wumpus if you are facing it
- Shooting uses up the only arrow
- Grabbing picks up the gold if in the same square
- Releasing drops the gold in the same square
Wumpus world characterization

Is the world deterministic??
Is the world fully accessible??
Is the world static??
Is the world discrete??

Wumpus world characterization

Is the world deterministic?? Yes—outcomes exactly specified
Is the world fully accessible?? No—only local perception
Is the world static?? Yes—Wumpus and Pits do not move
Is the world discrete?? Yes
Exploring a wumpus world

Other tight spots

Breeze in (1,2) and (2,1) 
⇒ no safe actions

Assuming pits uniformly distributed, 
(2,2) is most likely to have a pit

Smell in (1,1) 
⇒ cannot move

Can use a strategy of coercion:
- shoot straight ahead
- wumpus was there ⇒ dead ⇒ safe
- wumpus wasn’t there ⇒ safe
Wumpus Knowledge Base

Logic Symbols:
- $W_{ij}$ – Wumpus in cell $[i, j]$
- $B_{ij}$ – Breeze in cell $[i, j]$
- $G_{ij}$ – Glitter in cell $[i, j]$
- $P_{ij}$ – Pit in cell $[i, j]$
- $S_{ij}$ – Stench in cell $[i, j]$

Sample Rules:
- $R_1$: $\neg S_{1,1} \Rightarrow \neg W_{1,1} \land \neg W_{1,2} \land \neg W_{2,1}$
- $R_2$: $\neg S_{2,2} \Rightarrow \neg W_{1,1} \land \neg W_{1,2} \land \neg W_{2,1}$
- $R_3$: $\neg S_{1,2} \Rightarrow \neg W_{1,1} \land \neg W_{1,2} \land \neg W_{2,2} \land \neg W_{1,3}$

etc...
- $R_4$: $S_{1,2} \Rightarrow W_{1,1} \lor W_{1,2} \lor W_{1,1}$

Normal forms

Other approaches to inference use syntactic operations on sentences, often expressed in standardized forms.

Conjunctive Normal Form (CNF—universal)

- **clauses**
- **conjunction of disjunctions of literals**

- E.g., $(A \lor \neg B) \land (B \lor \neg C \lor \neg D)$

Disjunctive Normal Form (DNF—universal)

- **terms**
- **disjunction of conjunctions of literals**

- E.g., $(A \land B) \lor (A \land \neg C) \lor (A \land \neg D) \lor (\neg B \land \neg C) \lor (\neg B \land \neg D)$

Horn Form (restricted)

- **conjunction of Horn clauses** (clauses with $\leq 1$ positive literal)

- E.g., $(A \lor \neg B) \land (B \lor \neg C \lor \neg D)$

Often written as set of implications:

- $B \Rightarrow A$ and $(C \land D) \Rightarrow B$
Inference in Wumpus World

KB: \~S_{1,1} \~S_{2,1} \~S_{1,2} \~B_{1,1} B_{2,1} \~B_{1,2}

Infer: W_{1,3}

Truth Table: sound and complete

\[12 \text{ symbols} \Rightarrow 2^{12} = 4096 \text{ rows}\]

Inference rules: Use rules of prop. logic to infer the conclusion

Action in Wumpus World

- Rules in KB to help decide what to do
  - ex: \( A_{1,1} \wedge \text{East} \wedge W_{2,1} \Rightarrow \sim\text{Forward} \)
- Ask KB what action to take
- Prop logic can’t answer: “What action should I take?”
- Can only answer: “Should I go forward?”
- In general: Too many propositions to handle
  - “Don’t go forward if Wumpus in front of you” \(\Rightarrow\) 64 rules
- Can’t deal with change
  - Can’t keep track of where it has been
Another Example

- Problem 6.5 – *AIMA Textbook*

- If the unicorn is mythical, then it is immortal, but if it is not mythical, then it is a mortal mammal. If the unicorn is either immortal or a mammal, then it is horned. The unicorn is magical if it is horned.

- *Prove*: Unicorn is mythical.

- *Question*: Is it magical? Horned?
Propositional logic: Syntax

Propositional logic is the simplest logic—illustrates basic ideas

The proposition symbols \( P_1, P_2 \) etc are sentences

- If \( S \) is a sentence, \( \neg S \) is a sentence
- If \( S_1 \) and \( S_2 \) is a sentence, \( S_1 \land S_2 \) is a sentence
- If \( S_1 \) and \( S_2 \) is a sentence, \( S_1 \lor S_2 \) is a sentence
- If \( S_1 \) and \( S_2 \) is a sentence, \( S_1 \Rightarrow S_2 \) is a sentence
- If \( S_1 \) and \( S_2 \) is a sentence, \( S_1 \Leftrightarrow S_2 \) is a sentence
Propositional logic: Semantics

Each model specifies true/false for each proposition symbol

E.g. A B C

True True False

Rules for evaluating truth with respect to a model m:

\[ \neg S \text{ is true iff } S \text{ is false} \]
\[ S_1 \land S_2 \text{ is true iff } S_1 \text{ is true and } S_2 \text{ is true} \]
\[ S_1 \lor S_2 \text{ is true iff } S_1 \text{ is true or } S_2 \text{ is true} \]
\[ S_1 \Rightarrow S_2 \text{ is true iff } S_1 \text{ is false or } S_2 \text{ is true} \]
\[ \text{i.e., is false iff } S_1 \text{ is true and } S_2 \text{ is false} \]
\[ S_1 \Leftrightarrow S_2 \text{ is true iff } S_1 \Rightarrow S_2 \text{ is true and } S_2 \Rightarrow S_1 \text{ is true} \]

Models

Logicians typically think in terms of models, which are formally structured worlds with respect to which truth can be evaluated.

We say m is a model of a sentence \( \alpha \) if \( \alpha \) is true in m.

\[ M(\alpha) \] is the set of all models of \( \alpha \).

Then \( KB \models \alpha \) if and only if \( M(KB) \subseteq M(\alpha) \).

E.g. \( KB = \) Giants won and Reds won

\( \alpha = \) Giants won