Artificial Intelligence

Topic 11

Logical Agents

Reading: Russell and Norvig, Chapter 7, Sections 1–4
Outline

♦ Knowledge-based agents
♦ Wumpus world
♦ Logics
♦ Propositional (Boolean) logic
  – syntax
  – semantics
♦ Models
♦ Entailment
♦ Equivalence, validity, satisfiability
Knowledge bases

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

```
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(t))
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
Wumpus World PEAS description

Performance measure
gold +1000, death -1000
-1 per step, -10 for using the arrow

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

Actuators Left turn, Right turn,
Forward, Grab, Release, Shoot

Sensors Breeze, Glitter, Smell
Wumpus world characterization

Observable?? No—only local perception

Deterministic?? Yes—outcomes exactly specified

Static?? Yes—Wumpus and Pits do not move

Discrete?? Yes

Single-agent?? Yes—Wumpus is essentially a natural feature
Exploring a wumpus world (different world to previous picture)
Exploring a wumpus world
Exploring a wumpus world

A

B

P?

P?

OK

OK

A

A

Exploring a wumpus world
Exploring a wumpus world
Exploring a wumpus world
Exploring a wumpus world

[Diagram of a wumpus world grid with annotations]

Other tight spots

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

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
Logics

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; \( x^2 + y > y \) 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 \( x = 0, \ y = 6 \).
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 (negation [“not”])

If $S_1$ and $S_2$ are sentences, $S_1 \land S_2$ is a sentence (conjunction [“and”])

If $S_1$ and $S_2$ are sentences, $S_1 \lor S_2$ is a sentence (disjunction [“or”])

If $S_1$ and $S_2$ are sentences, $S_1 \Rightarrow S_2$ is a sentence (implication [“if-then”])

If $S_1$ and $S_2$ are sentences, $S_1 \iff S_2$ is a sentence (biconditional [“if-and-only-if”])
Let $P_{i,j}$ be true if there is a pit in $[i,j]$.
Let $B_{i,j}$ be true if there is a breeze in $[i,j]$.

$\neg P_{1,1}, \neg P_{1,2}, \neg P_{2,1}$
$\neg B_{1,1}, B_{1,2}, \neg B_{2,1}$
Wumpus world sentences

Let $P_{i,j}$ be true if there is a pit in $[i, j]$.
Let $B_{i,j}$ be true if there is a breeze in $[i, j]$.

$\neg P_{1,1}, \neg P_{1,2}, \neg P_{2,1}$
$\neg B_{1,1}, B_{1,2}, \neg B_{2,1}$

“Pits cause breezes in adjacent squares”

$P_{1,3} \Rightarrow (B_{1,2} \land B_{2,3} \land B_{1,4})$
$P_{2,2} \Rightarrow (B_{1,2} \land B_{2,3} \land B_{3,2} \land B_{2,1})$

Can we conclude $\neg P_{2,2}$, $P_{1,3}$?
Wumpus world sentences

Let $P_{i,j}$ be true if there is a pit in $[i, j]$.
Let $B_{i,j}$ be true if there is a breeze in $[i, j]$.

\[ \neg P_{1,1}, \neg P_{1,2}, \neg P_{2,1} \]
\[ \neg B_{1,1}, B_{1,2}, \neg B_{2,1} \]

“A square is breezy if and only if there is an adjacent pit”

\[ B_{1,2} \iff (P_{1,1} \lor P_{2,2} \lor P_{1,3}) \]
\[ B_{2,1} \iff (P_{1,1} \lor P_{2,2} \lor P_{3,1}) \]

Can we conclude $\neg P_{2,2}$, $P_{1,3}$?

What do we mean by “conclude”?
Propositional logic: Semantics

Specifies true/false for each proposition symbol

E.g. \( P_{1,2}, P_{2,2}, P_{3,1} \)

\[
\begin{array}{ccc}
\text{true} & \text{true} & \text{false}
\end{array}
\]

With \( n \) symbols, \( 2^n \) possible truth assignments can be enumerated automatically.

Rules for evaluating truth of compound sentences:

\[
\begin{align*}
\neg S & \text{ is true iff } S \text{ is false} \\
S_1 \land S_2 & \text{ is true iff } S_1 \text{ is true and } S_2 \text{ is true} \\
S_1 \lor S_2 & \text{ is true iff } S_1 \text{ is true or } S_2 \text{ is true} \\
S_1 \Rightarrow S_2 & \text{ is true iff } S_1 \text{ is false or } S_2 \text{ is true} \\
\text{i.e., is false iff } & S_1 \text{ is true and } S_2 \text{ is false} \\
S_1 \nleftrightarrow S_2 & \text{ is true iff } S_1 \Rightarrow S_2 \text{ is true and } S_2 \Rightarrow S_1 \text{ is true}
\end{align*}
\]
### Truth tables for connectives

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>¬P</th>
<th>P ∧ Q</th>
<th>P ∨ Q</th>
<th>P ⇒ Q</th>
<th>P⇔Q</th>
</tr>
</thead>
<tbody>
<tr>
<td>false</td>
<td>false</td>
<td>true</td>
<td>false</td>
<td>false</td>
<td>true</td>
<td>true</td>
</tr>
<tr>
<td>false</td>
<td>true</td>
<td>true</td>
<td>false</td>
<td>true</td>
<td>true</td>
<td>false</td>
</tr>
<tr>
<td>true</td>
<td>false</td>
<td>false</td>
<td>false</td>
<td>true</td>
<td>false</td>
<td>false</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>
</tbody>
</table>

Simple recursive process evaluates an arbitrary sentence, e.g.,
\[ \neg P_{1,2} \land (P_{2,2} \lor P_{3,1}) = true \land (false \lor true) = true \land true = true \]
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 \( \alpha \), or \( m \) satisfies \( \alpha \), if \( \alpha \) is true in \( m \).

For example, the model illustrated here satisfies the sentences:
\[ \alpha_1 = \text{there is a breeze in square [4,1]} \]
and:
\[ \alpha_2 = \text{every square adjacent to a pit has a breeze} \]
but not the sentence:
\[ \alpha_3 = \text{every square adjacent to a breeze has a pit} \]

We say a model \( m \) satisfies a set of sentences \( KB \) iff \( m \) satisfies all sentences \( \alpha \in KB \).
Entailment means that one thing follows from or is a logical consequence of another:

\[ KB \models \alpha \]

Knowledge base \( KB \) entails sentence \( \alpha \)
if and only if \( \alpha \) is true whenever \( KB \) is true

E.g., the KB containing “the Dockers won” and “the Eagles lost” entails “Either the Dockers won or the Eagles won” (irrespective of the other teams)

E.g., \( x + y = 4 \) entails \( 4 = x + y \)

Entailment is a relationship between sentences (i.e., syntax) that is based on semantics
Models and Entailment

Let $M(\alpha)$ be the set of all models satisfying $\alpha$, and $M(KB)$ be the set of all models satisfying $KB$.

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

E.g. $KB = \text{Dockers won and Eagles lost}$
\[ \alpha = \text{Dockers won} \]

Assume there are 4 teams in the league.

How many possible outcomes (assuming no draws)?

Which results satisfy $KB$? Which results satisfy $\alpha$?

Does $KB \models \alpha$?
Entailment in the wumpus world

Situation after detecting nothing in [1,1], moving right, breeze in [2,1]

Consider possible models for ?s assuming only pits

3 Boolean choices $\Rightarrow$ 8 possible models
Wumpus models
**Wumpus models**

\[ KB = \text{wumpus-world rules} + \text{observations} \]
\( KB = \text{wumpus-world rules} + \text{observations} \)

\( \alpha_1 = \text{“[1,2] is safe”}, \ KB \models \alpha_1, \text{proved by model checking} \)
$KB = \text{wumpus-world rules + observations}$
Wumpus models

\[ KB = \text{wumpus-world rules} + \text{observations} \]

\[ \alpha_2 = \text{“[2,2] is safe”}, \ KB \not\models \alpha_2 \]
Determining Entailment by Enumerating Truth Tables

Show that \( \{ \neg B_{2,1}, P_{2,2} \Rightarrow B_{1,2} \land B_{2,1} \} \models \neg P_{2,2} \)

<table>
<thead>
<tr>
<th>( B_{1,2} )</th>
<th>( B_{2,1} )</th>
<th>( P_{2,2} )</th>
<th>( \neg B_{2,1} )</th>
<th>( P_{2,2} \Rightarrow B_{1,2} \land B_{2,1} )</th>
<th>( \neg P_{2,2} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>F</td>
<td>F</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>T</td>
<td>T</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>F</td>
<td>F</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>T</td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>F</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>T</td>
<td>T</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>F</td>
<td>F</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>T</td>
<td>F</td>
<td>T</td>
<td>F</td>
</tr>
</tbody>
</table>

Enumerate rows (different assignments to symbols), if \( KB \) is true in row, check that \( \alpha \) is too.
Logical equivalence

Two sentences are logically equivalent iff true in same models:
\(\alpha \equiv \beta\) if and only if \(\alpha \models \beta\) and \(\beta \models \alpha\)

\[
\begin{align*}
(\alpha \land \beta) & \equiv (\beta \land \alpha) \quad \text{commutativity of } \land \\
(\alpha \lor \beta) & \equiv (\beta \lor \alpha) \quad \text{commutativity of } \lor \\
((\alpha \land \beta) \land \gamma) & \equiv (\alpha \land (\beta \land \gamma)) \quad \text{associativity of } \land \\
((\alpha \lor \beta) \lor \gamma) & \equiv (\alpha \lor (\beta \lor \gamma)) \quad \text{associativity of } \lor \\
\neg(\neg\alpha) & \equiv \alpha \quad \text{double-negation elimination} \\
(\alpha \Rightarrow \beta) & \equiv (\neg\beta \Rightarrow \neg\alpha) \quad \text{contraposition} \\
(\alpha \Rightarrow \beta) & \equiv (\neg\alpha \lor \beta) \quad \text{implication elimination} \\
(\alpha \iff \beta) & \equiv ((\alpha \Rightarrow \beta) \land (\beta \Rightarrow \alpha)) \quad \text{biconditional elimination} \\
\neg(\alpha \land \beta) & \equiv (\neg\alpha \lor \neg\beta) \quad \text{De Morgan} \\
\neg(\alpha \lor \beta) & \equiv (\neg\alpha \land \neg\beta) \quad \text{De Morgan} \\
(\alpha \land (\beta \lor \gamma)) & \equiv ((\alpha \land \beta) \lor (\alpha \land \gamma)) \quad \text{distributivity of } \land \text{ over } \lor \\
(\alpha \lor (\beta \land \gamma)) & \equiv ((\alpha \lor \beta) \land (\alpha \lor \gamma)) \quad \text{distributivity of } \lor \text{ over } \land
\end{align*}
\]
Validity and satisfiability

A sentence is **valid** if it is true in **all** models,

\[ \text{e.g., } \text{True}, \ A \lor \neg A, \ A \Rightarrow A, \ (A \land (A \Rightarrow B)) \Rightarrow B \]

A sentence is **satisfiable** if it is true in **some** model

\[ \text{e.g., } A \lor B, \ C \]

A sentence is **unsatisfiable** if it is true in **no** models

\[ \text{e.g., } A \land \neg A \]

Next — finding logical consequences...