Logical Agents

EECS 492
February 2nd, 2010

Mary Herchenhahn, mary-h.com

Administrative

- Thursday: Midterm 1, 7p-9p
  - DOW1013: Last name A-M
  - DOW1017: Last name N-Z

- Next Tuesday:
  - PS2 due
  - PS3 distributed--- new team assignments
Today

- An approach for dealing with incomplete knowledge

- Introduce an important problem domain: Wumpus

Incomplete Knowledge

- We don’t always know the actual world state

- Leads to two questions:
  - How do we represent incomplete knowledge?
  - How do we reason about incomplete knowledge?

- Previously mentioned: Belief state
  - Representation:
    - set of possible states
  - Reasoning:
    - our search methods from before.
Vacuum world: Belief States

Incomplete Knowledge

- Consider n room version of vacuum-world
  - Assume no dirt sensors
  - How many states?
    - $2^n$ possible world states
  - Search space: all subsets of world states

$$2^{2^n}$$

Much simpler (in this case): reasoning
Just go to every room and Suck!
Describing Knowledge

- Often can be much more concise than enumerating possible states
- Idea: construct and manipulate descriptions of knowledge
- Some descriptions implicit in others:
  - At least one room is dirty. Rm B is clean.
  - Is room A dirty?
  - How many rooms are dirty?

Hunt the Wumpus

- Let’s see how knowledge and reasoning can help us
- Today:
  - Focus on concepts and terminology
  - Use human intuition to solve problems
Wumpus world: rules

- **4x4 grid**
  - Pits (P)
    - Breeze (B)
  - Wumpus (W)
    - Stench (S)

Entering a pit is fatal. Can always sense a breeze from an adjoining room.

Encountering a wumpus is fatal. Can always sense a stench from an adjoining room.
Wumpus world: rules

- 4x4 grid
- Pits (P)
  - Breeze (B)
- Wumpus (W)
  - Stench (S)
- Gold (G)

We want to find the gold. We can only detect it when we’re on top of it.
Wumpus world: encoding

- Knowledge Representation:
  - $Z_{x,y}$ means "Z" at $(x,y)$
  - $\neg Z_{x,y}$ means "Z" not at $(x,y)$

- Wish to discover location of G and safe squares that let us get there!

Wumpus world: Example

- Percepts (Knowledge)
  - $A_{1,1}$
  - $\neg B_{1,1}$
  - $\neg S_{1,1}$

- Inference
  - $\neg P_{1,2}, \neg P_{2,1}$
  - $\neg W_{1,2}, \neg W_{2,1}$

- Rules of the game
Wumpus world: Example

Knowledge

\[ A_{1,1}, \neg B_{1,1}, \neg S_{1,1} \]
\[ \neg P_{1,2}, \neg P_{2,1} \]
\[ \neg W_{1,2}, \neg W_{2,1} \]
\[ S_{2,1}, \neg B_{2,1} \]
\[ \neg P_{2,2}, \neg P_{3,2} \]

Rules of the game

Inference

\[ \neg P_{2,2}, \neg P_{3,2} \]

Wumpus world: Example

Knowledge

\[ A_{1,1}, \neg B_{1,1}, \neg S_{1,1} \]
\[ \neg P_{1,2}, \neg P_{2,1} \]
\[ \neg W_{1,2}, \neg W_{2,1} \]
\[ S_{2,1}, \neg B_{2,1} \]
\[ S_{1,2}, \neg B_{1,2} \]

Rules of the game

Inference

\[ \neg P_{1,3}, W_{2,2} \]
What have we learned?

- We can do well in Hunt the Wumpus without enumerating belief states.
- We can reason about knowledge directly.

- But we (humans) were doing the reasoning! How do we get autonomous agents to do the reasoning?

Knowledge-Based Agent

![Diagram showing the interaction between Environment, Knowledge Base, Reasoning (Inference), and Agent. The percepts from the Environment feed into the Knowledge Base, which in turn queries the Reasoning (Inference) module with a question: "What should I do next?" The answer to this query is then acted upon by the Agent, forming a cycle between actions and percepts.]
Encoding knowledge

- Wumpus: we encoded *some* of our knowledge (the presence/absence of certain features)

- How do we encode the rules of the game?
  - e.g.: “The Wumpus emits a stench that can be detected from adjacent cells”

- We need a more powerful way of encoding knowledge!

Knowledge Representation

- Knowledge representation *language*: notation for expressing a KB

- Consists of
  - *Syntax*: defines the legal sentences
  - *Semantics*: *facts* in the world to which sentences correspond

- *Logic*: KR language with *well-defined* syntax and semantics
Sentence Syntax

Sentence \[ \rightarrow \text{AtomicSentence} \mid \text{ComplexSentence} \]

AtomicSentence \[ \rightarrow \text{True} \mid \text{False} \mid \text{Symbol} \]

Symbol \[ \rightarrow P \mid Q \mid R \mid \ldots \]

ComplexSentence \[ \rightarrow \neg \text{Sentence} \mid (\text{Sentence} \land \text{Sentence}) \mid (\text{Sentence} \lor \text{Sentence}) \mid (\text{Sentence} \Rightarrow \text{Sentence}) \mid (\text{Sentence} \Leftrightarrow \text{Sentence}) \]

BNF (Backus-Naur Form) grammar

Semantics

- Defines an interpretation for symbols in the logic
- Example:
  - Sentence “\(D_X\)” interpreted as fact that there is dirt in room \(X\).
  - Sentence is true if, in the real world, there actually is dirt in room \(X\).
- Model (aka possible world)
  - Specifies truth or falsity of every sentence
Logical Connectives

- Given boolean value(s), compute a new boolean value.

- Think: digital logic gates!

Logical Connectives: \( \neg \)

- Standard boolean “NOT”

<table>
<thead>
<tr>
<th>A</th>
<th>( \neg A )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
Logical Connectives: $\land$

- Standard boolean “AND”

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>A $\land$ B</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Logical Connectives: $\lor$

- Standard boolean “OR”
- Not Exclusive

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>A $\lor$ B</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Remembering $\land$ vs. $\lor$

- Helpful mnemonics:
  - $\land$ looks like “A” for “And”
  - $\lor$ looks like “V” for “Vel”

- Imagine rain falling from above:
  - $\lor$ collects more $\rightarrow$ OR
  - $\land$ collects less $\rightarrow$ AND

Logical Connectives: $\Rightarrow$

- Implication: $A \Rightarrow B$
- NOT the same as entailment, $A \models B$
- Note behavior when $\neg A$

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>$A \Rightarrow B$</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Logical Connectives: \( \Rightarrow \)

- Implication: \( A \Rightarrow B \)
- Equivalent expression?

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>A ( \Rightarrow B )</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Logical Connectives: \( \Leftrightarrow \)

- Biconditional \( A \Leftrightarrow B \)
  - True if \( A \Rightarrow B \) and \( B \Rightarrow A \)

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>A ( \Leftrightarrow B )</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Five Logical Connectives

<table>
<thead>
<tr>
<th>P</th>
<th>Q</th>
<th>\neg P (not)</th>
<th>P \land Q (and)</th>
<th>P \lor Q (or)</th>
<th>P \Rightarrow Q (implies)</th>
<th>P \Leftrightarrow Q (if and only if)</th>
</tr>
</thead>
<tbody>
<tr>
<td>false</td>
<td>false</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>false</td>
<td>true</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>true</td>
<td>false</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>true</td>
<td>true</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Return of the Wumpus

- **We can now encode the rules of the game!**

“The Wumpus emits a stench that can be detected from adjacent cells”

- \( S_{x,y} \Rightarrow W_{x-1,y} \lor W_{x+1,y} \lor W_{x,y-1} \lor W_{x,y+1} \) (right?)

- \( S_{x,y} \Leftrightarrow W_{x-1,y} \lor W_{x+1,y} \lor W_{x,y-1} \lor W_{x,y+1} \)
Order of Operations

- What does \( \neg A \land B \lor C \) mean?
  1. \( \neg ((A \land B) \lor C) \)
  2. \( \neg (A \land (B \lor C)) \)
  3. \( ((\neg A) \land B) \lor C \)
  4. \( (\neg (A \land B)) \lor C \)

- Always safe to use extra parentheses!

- Otherwise, order of operations:
  - Highest to lowest
  - \( \neg, \land, \lor \Rightarrow, \leftrightarrow \)

Properties of Sentences

- True or False:
  - Value of expression (with respect to a particular model)

- Valid:
  - True in all models

- Satisfiable:
  - True in some model

- Examples:
  - \( D_x \lor \neg D_x \)
  - \( D_x \lor D_y \)
  - \( D_x \land \neg D_x \)
Knowledge Base

- KB is the set of all known true sentences for the actual world model.
  - Contains generalizations ("game rules") applicable to all instances of "Hunt the Wumpus"
  - Contains percepts applicable to the particular instance we’re playing.

Entailment

- Suppose we want to know if $W_{2,2}$ is true, given KB. (does KB $\models W_{2,2}$?)
- We better make sure that there isn’t a model satisfying KB but where $W_{2,2}$ is false.
  - I.e., is $(KB \land \neg W_{2,2})$ unsatisfiable?
  - I.e., is $(\neg KB \lor W_{2,2})$ valid?
  - I.e., is $(KB \Rightarrow W_{2,2})$ valid?
- How is that different from the value of $(KB \Rightarrow W_{2,2})$?
Entailment

- Relation between sentences, says whether one is implicit in other(s).
  - KB: a set of sentences
  - $\alpha$: a sentence

$$\text{KB} \models \alpha$$

- KB entails $\alpha$
- In every model of KB (i.e., a model in which KB is true), $\alpha$ is true.
- Truth of $\alpha$ is contained in KB.

Deduction Theorem

- Entailment and Implication are different, but are related to each other via the Deduction Theorem.

- Deduction Theorem:
  - For any sentences A and B:
    $$((A \implies B) \text{ is valid}) \iff A \models B$$
Implication vs. Entailment

Entailment: For all world models in which A is true, B is true. (A => B is valid).

A ⇒ B

Implication: Compute a boolean value as a function of A and B equal to (¬A ∨ B). If ¬A, we’re not saying anything about B.

Inference

Process by which some sentences are derived from others.

Aka reasoning

Record of an inference process called a proof

Can derive α from KB using method i:

KB ⊢_i α
Properties of inference methods

- **Soundness**
  
  \[ \text{KB} \models^i \alpha \quad \text{only when} \quad \text{KB} \models \alpha \]

- **Completeness**

  whenever \( \text{KB} \models \alpha \) it is true that \( \text{KB} \models^i \alpha \)

---

Gottfried Wilhelm Leibnitz (1646-1716)

… if we could find characters or signs appropriate for expressing all our thoughts as definitely and as exactly as arithmetic expresses numbers..., we could in all subjects in so far as they are amenable to reasoning accomplish what is done in Arithmetic... For all inquiries... would be performed by the transposition of characters and by a kind of calculus, which would immediately facilitate the discovery of beautiful results...

—*Dissertio de Arte Combinatoria, 1666*
Inference Methods

- Now, all we need is an inference method we can implement!

Model Checking

- A generic inference mechanism
- Enumerate all models (i.e., truth assignments) and check that $\alpha$ is true in all models in which KB is true
  - Time complexity: $O(2^n)$
  - Space complexity: $O(n)$
- **Sound**:
  - Yes: directly implements the definition of entailment
- **Complete**
  - Yes: given finite KB and $\alpha$ (because there are only finitely many models to examine)
Model Checking Example

KB:
(IsDog(Fido) ∨ IsCat(Fido)) ∧
(IsCat(Fido) ⇔ Meows(Fido)) ∧
(¬ Meows(Fido))

Does KB |= IsDog(Fido)

<table>
<thead>
<tr>
<th>IsDog(Fido)</th>
<th>IsCat(Fido)</th>
<th>Meows(Fido)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Model Checking Example

KB:
(IsDog(Fido) ∨ IsCat(Fido)) ∧
(IsCat(Fido) ⇔ Meows(Fido)) ∧
(¬ Meows(Fido))

Does KB |= IsDog(Fido)

<table>
<thead>
<tr>
<th>IsDog(Fido)</th>
<th>IsCat(Fido)</th>
<th>Meows(Fido)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Model Checking Example

- **KB:**
  \[(\text{IsDog}(Fido) \lor \text{IsCat}(Fido)) \land \]
  \[(\text{IsCat}(Fido) \iff \text{Meows}(Fido)) \land \]
  \[\neg \text{Meows}(Fido)\]

- **Does** \(KB \models \text{IsDog}(Fido)\)

<table>
<thead>
<tr>
<th>IsDog(Fido)</th>
<th>IsCat(Fido)</th>
<th>Meows(Fido)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Model Checking Example

- **KB:**
  \[(\text{IsDog}(Fido) \lor \text{IsCat}(Fido)) \land \]
  \[(\text{IsCat}(Fido) \iff \text{Meows}(Fido)) \land \]
  \[\neg \text{Meows}(Fido)\]

- **Does** \(KB \models \text{IsDog}(Fido)\)

<table>
<thead>
<tr>
<th>IsDog(Fido)</th>
<th>IsCat(Fido)</th>
<th>Meows(Fido)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Model Checking Example

- **KB:**
  (IsDog(Fido) \lor IsCat(Fido)) \land
  (IsCat(Fido) \iff Meows(Fido)) \land
  (\neg Meows(Fido))

- **Does KB \models IsDog(Fido) ?**
  
  For all world models in which KB is true, IsDog(Fido) is true, so YES!

Next Time

- **Propositional Logic**
  - More ways to manipulate our knowledge
  - Inference without enumerating all states