Logical Agents

EECS 492
February 3rd, 2011

Administrative

- PS2 due today
- PS3 team assignments will occur at 11:59p tonight!
  - Update your team preferences!

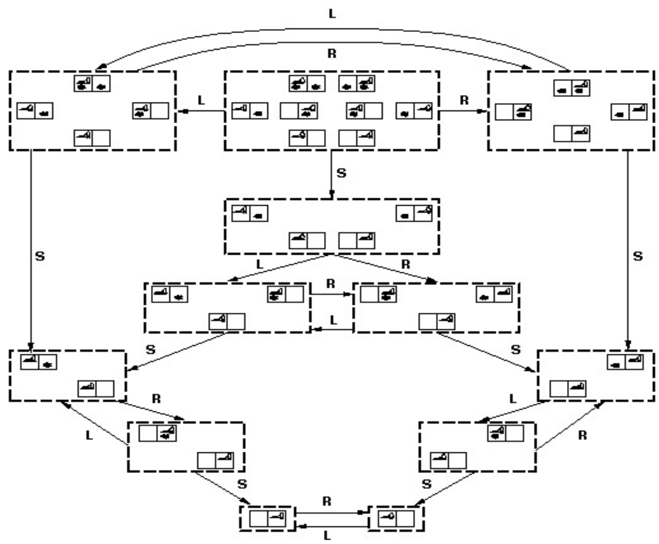
Today

- 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

$$2^{2^n}$$

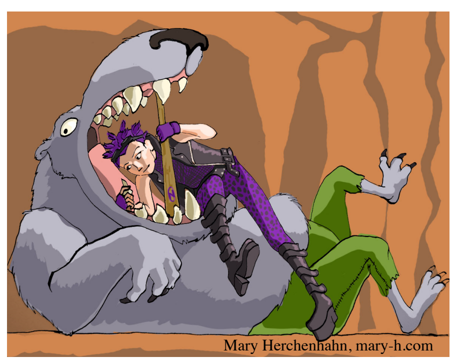
Much simpler (in this case): *reasoning*
Just go to every room and *Suck*!

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)

	B		
B	**P**	B	
	B		B
		B	**P**

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

Wumpus world: rules

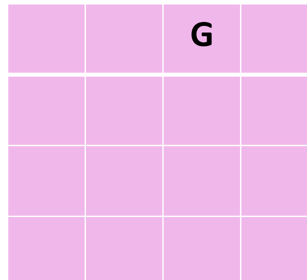
- 4x4 grid
  - Pits (P)
    - Breeze (B)
  - Wumpus (W)
    - Stench (S)

	S		
S	**W**	S	
	S		

Encountering a wumpus is fatal. Can always sense a stench from an adjoining room.

Wumpus world: rules

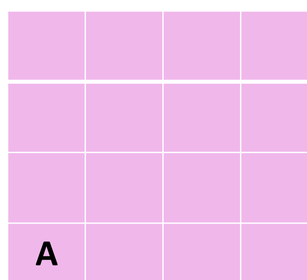
- 4x4 grid
  - Pits (P)
    - Breeze (B)
  - Wumpus (W)
    - Stench (S)
  - Gold (G)



We want to find the gold. We can only detect it when we're on top of it.

Wumpus world: rules

- 4x4 grid
  - Pits (P)
    - Breeze (B)
  - Wumpus (W)
    - Stench (S)
  - Gold (G)
  - Agent (A) starts at (1,1)



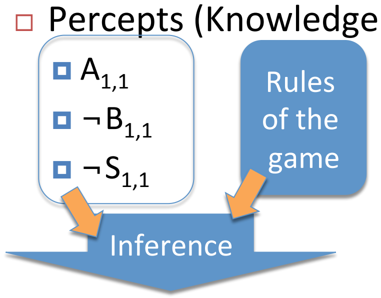
Wumpus world: encoding

- 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!

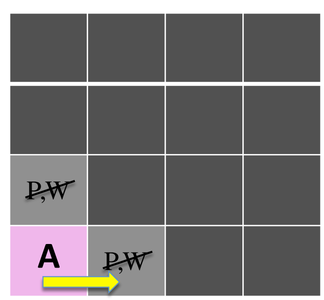
S	**W**	SB	**P**
	SB	**G**	B
B	**P**	B	
A	B		

Wumpus world: Example

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



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



Wumpus world: Example

- Knowledge

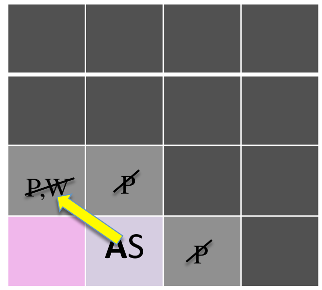
from before
- $A_{1,1}$, $\neg B_{1,1}$, $\neg S_{1,1}$
- $\neg P_{1,2}$, $\neg P_{2,1}$
- $\neg W_{1,2}$, $\neg W_{2,1}$

new
- $S_{2,1}$, $\neg B_{2,1}$

Rules of the game

Inference

- $\neg P_{2,2}$, $\neg P_{3,2}$



Wumpus world: Example

- Knowledge

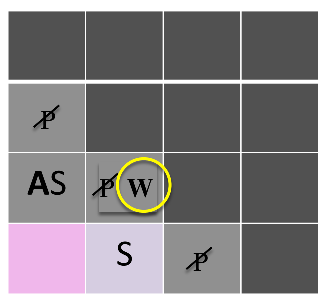
from before
- $A_{1,1}$, $\neg B_{1,1}$, $\neg S_{1,1}$
- $\neg P_{1,2}$, $\neg P_{2,1}$
- $\neg W_{1,2}$, $\neg W_{2,1}$
- $S_{2,1}$, $\neg B_{2,1}$
- $\neg P_{2,2}$, $\neg P_{3,2}$

new
- $S_{1,2}$, $\neg B_{1,2}$

Rules of the game

Inference

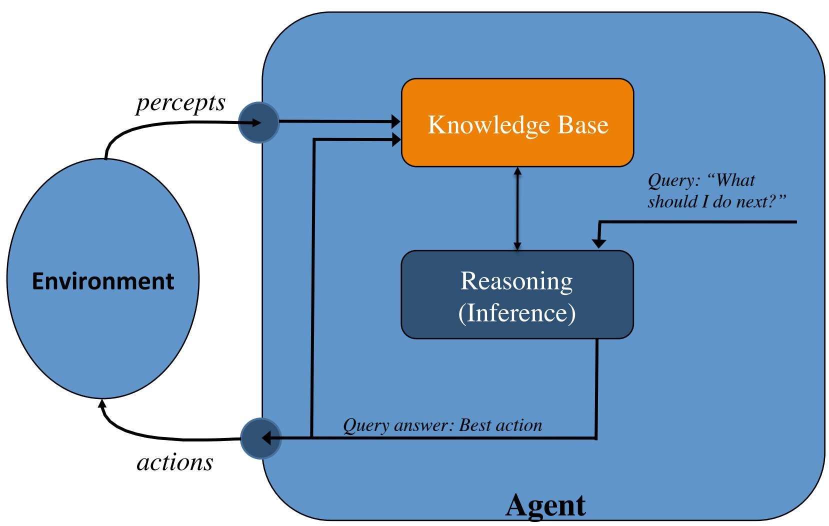
- $\neg P_{1,3}$, $W_{2,2}$



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



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 → AtomicSentence | ComplexSentence
AtomicSentence → **True** | **False** | Symbol
Symbol → **P** | **Q** | **R** | …
ComplexSentence → ¬Sentence
| (Sentence ∧ Sentence)
| (Sentence ∨ Sentence)
| (Sentence ⇒ Sentence)
| (Sentence ⇔ 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: ¬

- Standard boolean "NOT"

A	¬ A
0	**1**
1	**0**

Logical Connectives: ∧

- Standard boolean “AND”

A	B	A ∧ B
0	0	**0**
0	1	**0**
1	0	**0**
1	1	**1**

Logical Connectives: ∨

- Standard boolean “OR”
- Not Exclusive

A	B	A ∨ B
0	0	**0**
0	1	**1**
1	0	**1**
1	1	**1**

Remembering ∧ vs. ∨

- Helpful mnemonics:
- ∧ looks like "A" for "And"
- ∨ looks like "V" for "Vel"
- Imagine rain falling from above:
  - ∨ collects more → OR
  - ∧ collects less → AND

Logical Connectives: ⇒

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

A	B	A ⇒ B
0	0	**1**
0	1	**1**
1	0	**0**
1	1	**1**

Logical Connectives: $\Rightarrow$

- Implication: $A \Rightarrow B$
- Equivalent expression?

A	B	$A \Rightarrow B$
0	0	**1**
0	1	**1**
1	0	**0**
1	1	**1**

Logical Connectives: $\Leftrightarrow$

- Biconditional $A \Leftrightarrow B$
  - True if $A \Rightarrow B$ and $B \Rightarrow A$

A	B	$A \Leftrightarrow B$
0	0	**1**
0	1	**0**
1	0	**0**
1	1	**1**

Five Logical Connectives

P	Q	¬P (not)	P∧Q (and)	P∨Q (or)	P⇒Q (implies)	P⇔Q (if and only if)
false	false	1	0	0	1	1
false	true	1	0	1	1	0
true	false	0	0	1	0	0
true	true	0	1	1	1	1

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"

- ~~$S_{x,y} \Rightarrow W_{x-1,y} \vee W_{x+1,y} \vee W_{x,y-1} \vee W_{x,y+1}$~~

(right?)

- $S_{x,y} \Leftrightarrow W_{x-1,y} \vee W_{x+1,y} \vee W_{x,y-1} \vee W_{x,y+1}$

Order of Operations

- What does ¬A∧B∨C mean?
  - 1. ¬((A∧B)∨C)
  - 2. ¬(A∧(B∨C))
  - 3. ((¬ A)∧B)∨C
  - 4. (¬(A∧B))∨C
- Always safe to use extra parentheses!
- Otherwise, order of operations:
  - Highest to lowest
  - ¬, ∧, ∨, ⇒, ⇔

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:
  - $D_X \vee \neg D_X$
  - $D_X \vee D_Y$
  - $D_X \wedge \neg D_X$

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 $(KB \wedge (\neg W_{2,2}))$ unsatisfiable?
  - I.e., is $(\neg KB \vee W_{2,2})$ valid?
  - I.e., is $(KB \Rightarrow W_{2,2})$ valid?
- How is that different from the value of $(KB \Rightarrow W_{2,2})$?

A	B	$A \Rightarrow B$
0	0	1
0	1	1
1	0	0
1	1	1

Entailment

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

$$KB \models \alpha$$

- KB entails α
- In every model of KB (i.e., a model in which KB is true), α is true.
- Truth of α 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 $\Rightarrow$ B) is valid) iff

$$A \models B$$

Implication vs. Entailment

$$A \models B$$

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

$$A \Rightarrow B$$

- Implication: Compute a boolean value as a function of A and B equal to ($\neg A \vee B$). If $\neg A$, we're not saying anything about B.

Next Time

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