Knowledge-Based Agents Chapter 7.1-7.3

Some material adopted from notes by Andreas Geyer-Schulz and Chuck Dyer

Big Idea

- Drawing reasonable conclusions from a set of data (observations, beliefs, etc.) seems key to intelligence
- Logic is a powerful and well developed approach to this and highly regarded by people
- Logic is also a strong formal system that we can programs computers to use
- Maybe we can reduce any AI problem to figuring out how to represent it in logic and apply standard proof techniques to generate solutions

Inference in People

- People can do logical inference, but are not very good at it
- Reasoning with negation and disjunction seems to be particularly difficult
- But, people seem to employ many kinds of reasoning strategies, most of which are neither complete nor *sound*

Wason Selection Task

- I have a pack of cards each of which has a letter written on one side and a number written on the other side.
- I claim the following rule is true:

If a card has a vowel on one side, then it has an even number on the other side.

• Look at these cards and say which card or cards to turn over in order to decide whether the rule is true or false?



Wason Selection Task

• Wason (1966) showed that people are not very good at this task.

Wason, P. (1966). Reasoning. In New Horizons in Psychology. Penguin, Hammondsworth, UK.

- To disprove P=>Q, find a situation in which $P \land \neg Q$
- To disprove vowel => even, find a card with a vowel and an odd number
- Thus, turn over the cards showing vowels and turn over cards showing odd numbers

Wason Selection Task

- This version seems easier for people to do, as was shown by Griggs & Cox, 1982
- Your are the bouncer in a bar. Which of these people do you card given the rule:

You must be 21 or older to drink beer.



• It may be simpler because it's more familiar or because people have special strategies to reason about certain situations, such as "cheating" in a social situation

Negation in Natural Language

- We often model the meaning of a natural language sentence as a logic statement
- This maps the following into equivalent statements
 - -All elephants are gray
 - -No elephant are not gray
- Double negation is common in informal language
 That won't do you no good
- But what does this mean
 - -We cannot underestimate the importance of logic

Logic as a Methodology

Even if people do not use formal logical reasoning for solving a problem, logic might be a good approach for AI for a number of reasons

- -Airplanes don't need to flap their wings
- -Logic may be a good implementation strategy
- Developing a solution in a formal system like logic can offer other benefits, e.g., letting us prove properties of the approach
- •See neats vs. scruffies

A knowledge-based agent

- A knowledge-based agent includes a knowledge base and an inference system
- A knowledge base is a set of representations of facts of the world
- Each individual representation is called a **sentence**
- The sentences are expressed in a knowledge representation language
- The agent operates as follows:
 - 1. It TELLs the knowledge base what it perceives
 - 2. It ASKs the knowledge base what action it should perform
 - 3. It performs the chosen action

Architecture of a KB agent

- Knowledge Level
 - The most abstract level: describe agent by saying what it knows
 - Ex: A taxi agent might know that the Golden Gate Bridge connects San Francisco with the Marin County

• Logical Level

- The level at which the knowledge is encoded into *sentences*
- Ex: links(GoldenGateBridge, SanFrancisco, MarinCounty)
- Implementation Level
 - Physical representation of the sentences in the logical level
 - -Ex: as a tuple serialized as (links goldengatebridge sanfrancisco marincounty)



The Wumpus World environment

- The Wumpus computer game
- The agent explores a cave consisting of rooms connected by passageways
- Lurking somewhere in the cave is the Wumpus, a beast that eats any agent that enters its room
- Some rooms contain bottomless pits that trap any agent that wanders into the room
- The Wumpus can fall into a pit too, so avoids them
- Occasionally, there is a heap of gold in a room.
- The goal is to collect the gold and exit the world without being eaten

Jargon file on "Hunt the Wumpus"

WUMPUS /wuhm'p*s/ n. The central monster (and, in many versions, the name) of a famous family of very early computer games called "Hunt The Wumpus," dating back at least to 1972 (several years before ADVENT) on the Dartmouth Time-Sharing System. The wumpus lived somewhere in a cave with the topology of a dodecahedron's edge/vertex graph (later versions supported other topologies, including an icosahedron and Mobius strip). The player started somewhere at random in the cave with five "crooked arrows"; these could be shot through up to three connected rooms, and would kill the wumpus on a hit (later versions introduced the wounded wumpus, which got very angry). Unfortunately for players, the movement necessary to map the maze was made hazardous not merely by the wumpus (which would eat you if you stepped on him) but also by bottomless pits and colonies of super bats that would pick you up and drop you at a random location (later versions added "anaerobic termites" that ate arrows, bat migrations, and earthquakes that randomly changed pit locations).

This game appears to have been the first to use a non-random graphstructured map (as opposed to a rectangular grid like the even older Star Trek games). In this respect, as in the dungeon-like setting and its terse, amusing messages, it prefigured ADVENT and Zork and was directly ancestral to both. (Zork acknowledged this heritage by including a super-bat colony.) Today, a port is distributed with SunOS and as freeware for the Mac. A C emulation of the original Basic game is in circulation as freeware on the net.

Wumpus History

• See <u>Hunt the Wumpus</u> for details



- Early (c. 1972) text-based game written in BASIC written by Gregory Yob, a student at UMASS, Dartmouth
- Defined a genre of games including adventure, zork, and nethack
- Eventually commercialized (c. 1980) for early personal computers
- The <u>Hunt the Wumpus basic code</u> is available in a 1976 article in Creative Computing by Yob!

AIMA's Wumpus World

4

3

The agent always starts in the field [1,1]

Agent's task is to find the gold, return ² to the field [1,1] and climb out of the cave



Agent in a Wumpus world: Percepts

- The agent perceives
 - stench in the square containing the Wumpus and in the adjacent squares (not diagonally)
 - breeze in the squares adjacent to a pit
 - glitter in the square where the gold is
 - **bump**, if it walks into a wall
 - awoeful **scream** everywhere in the cave, if Wumpus is killed
- Percepts given as five-symbol list, e.g., if stench and breeze, but no glitter, bump or scream: [Stench, Breeze, None, None, None]
- Agent cannot perceive its own location

Wumpus World Actions

- go forward
- turn right 90 degrees
- turn left 90 degrees
- grab: Pick up object in the same square as the agent
- **shoot**: Fire arrow in straight line in direction agent is facing. It continues until it hits and kills Wumpus or hits outer wall. The agent has only one arrow, so only first shoot action has an effect
- climb is used to leave cave, only effective in start square
- **die**: This action automatically and irretrievably happens if agent enters square with pit or live Wumpus

Wumpus World Goal

The agent's goal is to find the gold and bring it back to the start square as quickly as possible, without getting killed

- -1,000 point reward for climbing out of the cave with the gold
- -1 point deducted for every action taken
- -10,000 point penalty for getting killed

Wumpus world characterization

- Fully Observable
- Deterministic
- Episodic
- Static
- Discrete
- Single-agent

Wumpus world characterization

- Fully Observable No only local perception
- **Deterministic** Yes outcomes exactly specified
- **Episodic** No sequential at the level of actions
- **Static** Yes Wumpus and Pits do not move
- **Discrete** Yes
- **Single-agent?** Yes Wumpus is essentially a natural feature

AIMA's Wumpus World

4

3

The agent always starts in the field [1,1]

Agent's task is to find the gold, return to the field [1,1] and climb out of the cave



The Hunter's first step

= Agent = Breeze

Р

S

1,4	2,4	3,4	4,4
1,3	2,3	3,3	4,3
1,2 OK	2,2	3,2	4,2
1,1 А ОК	2,1 OK	3,1	4,1

(a)

Since the agent is alive and perceives neither a breeze nor a stench at [1,1], it knows that [1,1] and its neighbors are OK

Α	= Agent	1,4	2,4	3,4	4,4
В	= Breeze				
G	= Glitter, Gold				
OK	= Safe square				
Р	= Pit	1,3	2,3	3.3	4,3
S	= Stench	,	,		,
V	= Visited				
W	= Wumpus				
		1,2	2.2 P?	3,2	4,2
		ок	$\neg \mathbf{W}$		
		1,1	^{2,1} A	^{3,1} P?	4,1
		V	В	⊐W	
		ок	ок	••	
			(t)	

Moving to [2,1] is a safe move that reveals a breeze but no stench, implying that the Wumpus is not adjacent but that one or more pits are

ок		
ок А	ок	

A	agent
В	breeze
G	glitter
OK	safe cell
Ρ	pit
S	stench
W	wumpus

в [ок А		
[ок А	ок	

A	agent
В	breeze
G	glitter
OK	safe cell
Р	pit
S	stench
W	wumpus



A	agent
В	breeze
G	glitter
OK	safe cell
Р	pit
S	stench
W	wumpus



A	agent
В	breeze
G	glitter
OK	safe cell
Ρ	pit
S	stench
W	wumpus



A	agent
В	breeze
G	glitter
OK	safe cell
Ρ	pit
S	stench
W	wumpus



agent
breeze
glitter
safe cell
pit
stench
wumpus



A	agent
В	breeze
G	glitter
OK	safe cell
Р	pit
S	stench
W	wumpus



А	agent
В	breeze
G	glitter
OK	safe cell
Ρ	pit
S	stench
W	wumpus



А	agent
В	breeze
G	glitter
OK	safe cell
Ρ	pit
S	stench
W	wumpus

Wumpuses online

- AIMA code
 - -<u>Python</u>
 - -<u>Lisp</u>
- <u>http://scv.bu.edu/cgi-bin/wcl</u> Web-based version you can play
- <u>http://codenautics.com/wumpus/</u> Mac version

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
 - $x+2 \ge y$ is a sentence; $x^2+y \ge \{\}$ is not a sentence
 - $x+2 \ge y$ is true iff the number x+2 is no less than the number y
 - $x+2 \ge y$ is true in a world where x = 7, y = 1
 - $x+2 \ge y$ is false in a world where x = 0, y = 6
 - x+1 > x is true for all numbers x

Entailment

- Entailment means that one thing follows from another:
- KB |= α
- Knowledge base *KB* entails sentence α iff α is true in *all possible worlds* where *KB* is true
 - E.g., the KB containing "UMBC won" and "JHU won" entails "Either UMBC won or JHU won"
 - -E.g., x+y = 4 entails 4 = x+y
 - Entailment is a relationship between sentences
 (i.e., syntax) that is based on semantics

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 α if α is true in *m*
- $M(\alpha)$ is the set of all models of α
- Then KB $\models \alpha$ iff $M(KB) \subseteq M(\alpha)$ -KB = UMBC and JHU won
 - $-\alpha = UMBC$ won
 - -Then 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 *KB* assuming only pits and restricting cells to {(1,3)(2,1)(2,2)}
- Two observations: ~B11, B12
- Three propositional variables variables: P13, P21, P22
- \Rightarrow 8 possible models

?	?	

P13	P21	P22
F	F	F
F	F	Т
F	Т	F
F	Т	Т
Т	F	F
Т	F	Т
Т	Т	F
Т	Т	Т

















3

STORE :

z

1

1

Wumpus World Rules (1)

- If a cell has a pit, then a breeze is observable in every adjacent cell
- In propositional calculus we can not have rules with variables (e.g., forall X...)

$$P11 \implies B21$$

 $P11 \implies B12$
 $P21 \implies B11$
 $P21 \implies B22$

these also follow $\sim B21 \Rightarrow \sim P11$ $\sim B12 \Rightarrow \sim P11$ $\sim B11 \Rightarrow \sim P21$ $\sim B22 \Rightarrow \sim P21$



KB = wumpus-world rules + observations

Wumpus World Rules (2)

- A cell is safe if it has neither a pit or wumpus $OK11 \Longrightarrow \sim P11 \land \sim W11$ $OK12 \Longrightarrow \sim P12 \land \sim W12$
- From which we can derive
 P11 V W11 => ~OK11
 P11 => ~OK11
 W11 => ~OK11

. . .



- *KB* = wumpus-world rules + observations
- $\alpha_1 = "[1,2]$ is safe"
- Since all models include α_1
- $KB \models \alpha_1$, proved by model checking



• *KB* = wumpus-world rules + observations



- *KB* = wumpus-world rules + observations
- $\alpha_2 = "[2,2]$ is safe"
- Since there are some models that don't include α_2
- KB $\not\models \alpha_2$
- We can not prove OK22. It might be true or false.

Inference, Soundness, Completeness

- $KB \models_i \alpha =$ sentence α can be derived from KB by procedure *i*
- Soundness: *i* is sound if whenever $KB \models_i \alpha$, it is also true that $KB \models \alpha$
- **Completeness:** *i* is complete if whenever $KB \models \alpha$, it is also true that $KB \models_i \alpha$
- 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*.

Representation, reasoning, and logic

- The object of knowledge representation is to express knowledge in a **computer-tractable** form, so that agents can perform well
- A knowledge representation language is defined by:
 - its syntax, which defines all possible sequences of symbols that constitute sentences of the language.
 - Ex: Sentences in a book, bit patterns in computer memory
 - its semantics, which determines the facts in the world to which the sentences refer.
 - Each sentence makes a claim about the world.
 - An agent is said to believe a sentence about the world.

The connection between sentences and facts



Semantics maps sentences in logic to facts in the world. The property of one fact following from another is mirrored by the property of one sentence being entailed by another.

Soundness and completeness

- A *sound* inference method derives only entailed sentences
- Analogous to the property of *completeness* in search, a *complete* inference method can derive any sentence that is entailed

Logic as a KR language



Ontology and epistemology

- **Ontology** is the study of what there is—an inventory of what exists. An ontological commitment is a commitment to an existence claim.
- **Epistemology** is a major branch of philosophy that concerns the forms, nature, and preconditions of knowledge.

Language	Ontological Commitment (What exists in the world)	Epistemological Commitment (What an agent believes about facts)
Propositional logic	facts	true/false/unknown
First-order logic	facts, objects, relations	true/false/unknown
Temporal logic	facts, objects, relations, times	true/false/unknown
Probability theory	facts	degree of belief 01
Fuzzy logic	degree of truth	degree of belief 01

No independent access to the world

- The reasoning agent often gets its knowledge about the facts of the world as a sequence of logical sentences and must draw conclusions only from them without independent access to the world.
- Thus, it is very important that the agent's reasoning is sound!



Summary

- Intelligent agents need knowledge about the world for making good decisions.
- The knowledge of an agent is stored in a knowledge base (KB) in the form of **sentences** in a knowledge representation (KR) language.
- A knowledge-based agent needs a **KB** and an **inference mechanism**. It operates by storing sentences in its knowledge base, inferring new sentences with the inference mechanism, and using them to deduce which actions to take.
- A **representation language** is defined by its syntax and semantics, which specify the structure of sentences and how they relate to the facts of the world.
- The **interpretation** of a sentence is the fact to which it refers. If the fact is part of the actual world, then the sentence is true.