

CMSC 471

Fall 2012

Class #10

Tuesday, October 2
Knowledge-Based Agents

Kevin Winner, winnerk1@umbc.edu

Knowledge-Based Agents

Chapter 7.1-7.3

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

Today's Class

- Knowledge-Based Agents
 - Representation
 - Wumpus World
 - Entailment and Derivation
- Project Description

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 and TELLS the kb about its action.

Architecture of a Knowledge-Based Agent

- **Knowledge Level.**

- The most abstract level: describe agent by saying what it knows.
- Example: 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.
- Example: `Links(GoldenGateBridge, SanFrancisco, MarinCounty)`.

- **Implementation Level.**

- The physical representation of the sentences in the logical level.
- Example: ``(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.
- Occasionally, there is a heap of gold in a room.
- The goal is to collect the gold and exit the world without being eaten

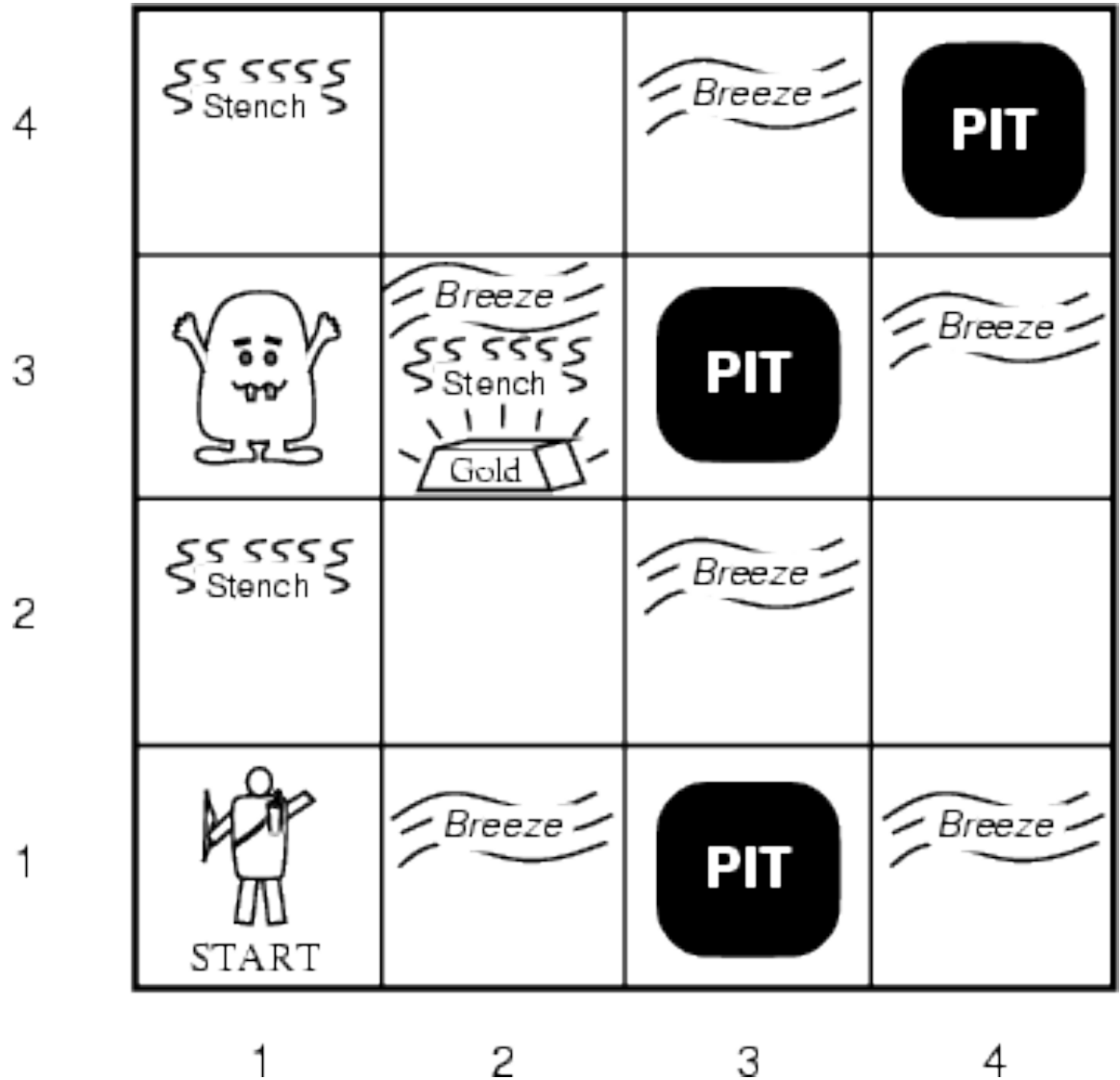
The Wumpus World Environment

- 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.
- Occasionally, there is a heap of gold in a room.
- The goal is to collect the gold and exit the world without being eaten



A Typical Wumpus World

- The agent usually starts in the field [1,1].
- The task of the agent is to find the gold, return to the field [1,1] and climb out of the cave.



Agent in a Wumpus World: Percepts

- The agent perceives
 - a stench in the square containing the wumpus and in the adjacent squares (not diagonally)
 - a breeze in the squares adjacent to a pit
 - a glitter in the square where the gold is
 - a bump, if it walks into a wall
 - a woeful scream everywhere in the cave, if the wumpus is killed
- The percepts are given as a five-symbol list. If there is a stench and a breeze, but no glitter, no bump, and no scream, the percept is
[Stench, Breeze, None, None, None]
- The agent cannot perceive its own location

Wumpus Actions

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

Wumpus 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

- 1000 points reward for climbing out of the cave with the gold
- 1 point deducted for every action taken
- 10000 points penalty for getting killed

The Wumpus Agent's First Step

1,4	2,4	3,4	4,4
1,3	2,3	3,3	4,3
1,2	2,2	3,2	4,2
OK			
1,1	2,1	3,1	4,1
A			
OK	OK		

(a)

- A** = Agent
- B = Breeze
- G = Glitter, Gold
- OK = Safe square
- P = Pit
- S = Stench
- V = Visited
- W = Wumpus

1,4	2,4	3,4	4,4
1,3	2,3	3,3	4,3
1,2	2,2	3,2	4,2
OK	P? -W		
1,1	2,1	3,1	4,1
V	A	P?	
OK	B	-W	
	OK		

(b)

Later

1,4	2,4	3,4	4,4
1,3 W!	2,3	3,3	4,3
1,2 A S OK	2,2 -W -P OK	3,2	4,2
1,1 V OK	2,1 B V OK	3,1 P! -W	4,1

(a)

A = Agent
B = Breeze
G = Glitter, Gold
OK = Safe square
P = Pit
S = Stench
V = Visited
W = Wumpus

1,4	2,4 P?	3,4	4,4
1,3 W!	2,3 A S G B	3,3 P?	4,3
1,2 S V OK	2,2 -W V -P OK	3,2	4,2
1,1 V OK	2,1 B V OK	3,1 P! -W	4,1

(b)

Models and Consistency

- A **model** of the world is an instantiation of all of the variables which compose the world
 - Typically these variables are boolean properties
 - Models are hypothetical belief states about the environment
- A model M is **consistent** with a knowledge base KB if all the sentences in KB are true in M

Entailment and Derivation

- **Entailment: $KB \models Q$**

- Q is entailed by KB (a set of premises or assumptions) if and only if there is no logically possible world in which Q is false while all the premises in KB are true.
- Or, stated positively, Q is entailed by KB if and only if the conclusion is true in every logically possible world in which all the premises in KB are true.

- **Derivation: $KB \vdash Q$**

- We can derive Q from KB if there is a proof consisting of a sequence of valid inference steps starting from the premises in KB and resulting in Q

KB Agents - Summary

- Intelligent agents need knowledge about the world for making good decisions.
- The knowledge of an agent is stored in a knowledge base in the form of **sentences** in a knowledge representation language.
- A knowledge-based agent needs a **knowledge base** 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 this fact is part of the actual world, then the sentence is true.

Mastermind