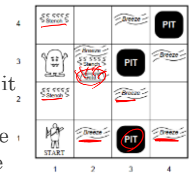
Logical Agents

ARTIFICIAL INTELLIGENCE
JUCHEOL MOON

1

Wumpus World description

- 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

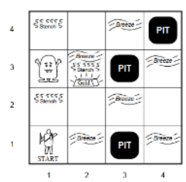


2

2

Wumpus World description

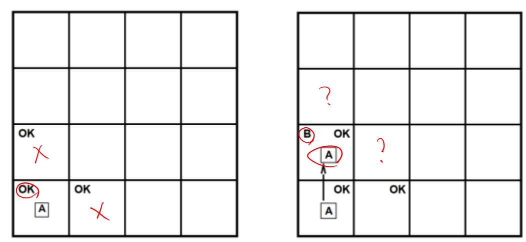
- Performance measure
  - gold +1000, death -1000
  - -1 per step, -10 for using the arrow
- Actuators
  - Up, Down, Left, Right move
    - No diagonal movement
  - Grab, Release, Shoot
- Sensors
  - Breeze, Glitter, Smell
- http://thiagodnf.github.io/wumpus-world-simulator/



3

3

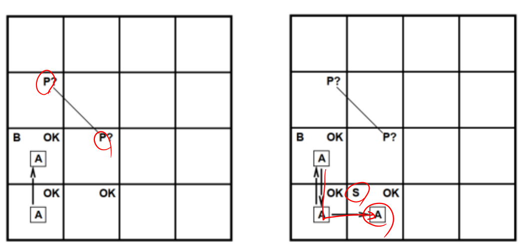
Exploring a wumpus world



4

4

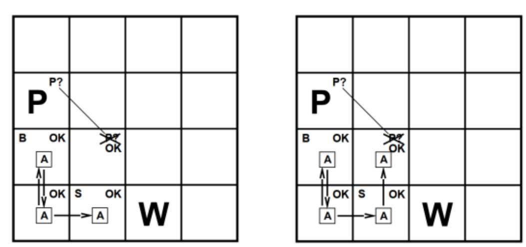
Exploring a wumpus world



5

5

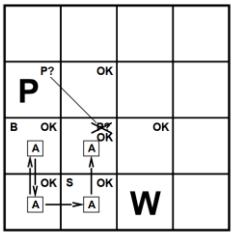
Exploring a wumpus world

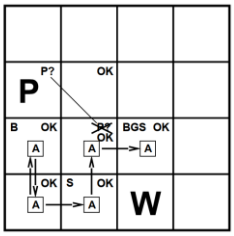


6

6

Exploring a wumpus world



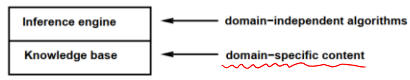


7

7

Knowledge bases

- Knowledge base
  - set of __setences__ in a formal language
- Declarative approach to building an agent
  - Tell it what it needs to know
- Then it can Ask itself what to do
  - answers should follow from the KB



8

8

The language of arithmetic

- Logics are formal languages for representing information such that conclusions can be drawn
- $x + 2 \geq y$ is a sentence
- $x2 + y >$ is not a sentence
- $x + 2 \geq y$ is true / false
  - in a world where $x = 7; y = 1$
- $x + 2 \geq y$ is true / false
  - in a world where $x = 0; y = 6$

9

9

Entailment

- Entailment means that one thing follows from another:
  - $KB \vDash \alpha$
  - Knowledge base KB entails sentence α if and only if α is **true** in all worlds where KB is **true**
  - E.g., $x + y = 4$ entails $4 = x + y$

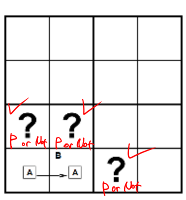
KB

10

10

Entailment in the wumpus world

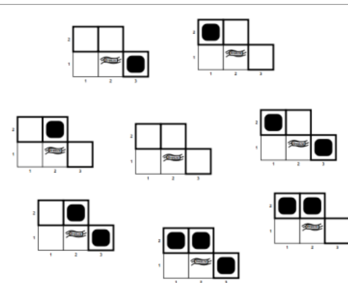
- Situation after detecting nothing in [1,1],
  - moving right, breeze in [2,1]
- Consider possible models for ?s
  - In terms of Pit
  - Ignore observations
  - They do not need be true
- 3 Boolean choices
  - ➔ __2^3__ possible models



11

11

Wumpus models

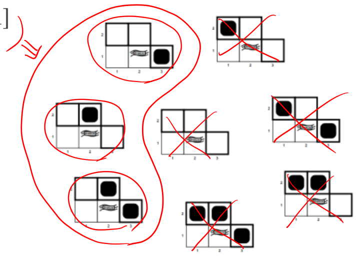


12

12

Wumpus models

- KB = wumpus-world rules + observations
  - nothing in [1,1]
  - breeze in [2,1]



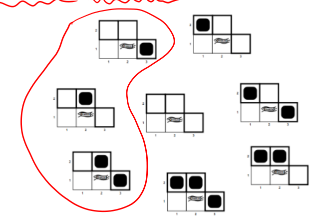
13

13

Wumpus models

$KB \vDash \alpha$ if and only if
α is **true** in all worlds where KB is **true**

- KB = wumpus-world rules + observations
- α_1 = "[1,2] is safe", $KB \vDash \alpha_1$? Yes



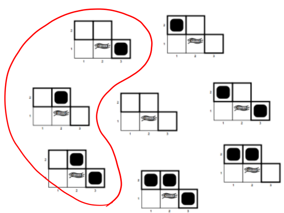
14

14

Wumpus models

$KB \vDash \alpha$ if and only if
α is **true** in all worlds where KB is **true**

- KB = wumpus-world rules + observations
- α_2 = "[2,2] is safe", $KB \vDash \alpha_2$? No

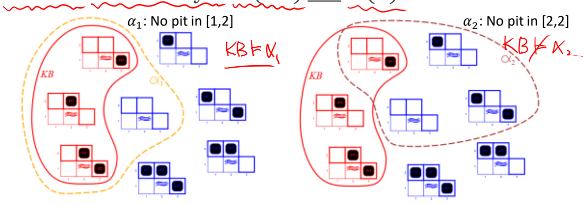


15

15

Models

- 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
  - $KB \vDash \alpha$ if and only if $M(KB) \subseteq M(\alpha)$



16

Propositional logic: Syntax

- Propositional logic is the simplest logic
  - illustrates basic ideas
- The proposition symbols P_1, P_2 are sentences
- If P is a sentence, $\neg P$ is a sentence
  - negation

17

17

Propositional logic: Syntax

- If P and Q are sentences, $P \wedge Q$ is a sentence
  - conjunction
- If P and Q are sentences, $P \vee Q$ is a sentence
  - disjunction
- If P and Q are sentences, $P \Rightarrow Q$ is a sentence
  - implication
- If P and Q are sentences, $P \Leftrightarrow Q$ is a sentence
  - biconditional

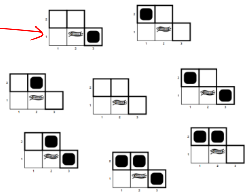
18

18

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 $[i, j]$.

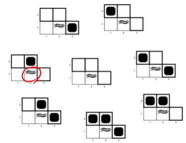
$\neg P_{1,1} = true$
$B_{2,1} = true$
$P_{3,1} = true$



19

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 $[i, j]$.
- Pits cause breezes in adjacent squares
  - A square is breezy if and only if there is an adjacent pit.
  - $B_{2,1} \Leftrightarrow P_{3,1} \vee P_{2,2} \vee P_{1,1}$
  - $B_{1,1} \Leftrightarrow P_{1,2} \vee P_{2,1}$



20

20

Satisfiability

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

$(x \vee y)$
$(x \vee y) \wedge z$

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

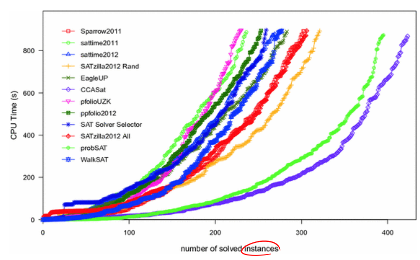
$x \wedge \neg x$

23

23

Satisfiability

- Is $B_{1,1} \Leftrightarrow (P_{1,2} \vee P_{2,1})$ satisfiable?



24

24

Conversion to CNF

Conjunctive Normal Form

- Is $B_{1,1} \Leftrightarrow (P_{1,2} \vee P_{2,1})$ satisfiable?

1. Replacing $\alpha \Leftrightarrow \beta$ with $(\alpha \Rightarrow \beta) \wedge (\beta \Rightarrow \alpha)$

$[B_{1,1} \Rightarrow (P_{1,2} \vee P_{2,1})] \wedge [(P_{1,2} \vee P_{2,1}) \Rightarrow B_{1,1}]$

2. Replacing $\alpha \Rightarrow \beta$ with $\neg\alpha \vee \beta$

$[\neg B_{1,1} \vee P_{1,2} \vee P_{2,1}] \wedge [\neg(P_{1,2} \vee P_{2,1}) \vee B_{1,1}]$

25

25

Conversion to CNF

- Is $B_{1,1} \Leftrightarrow (P_{1,2} \vee P_{2,1})$ satisfiable?

3. Move $\neg$ inwards using de Morgan's rules and double-negation

$(\neg B_{1,1} \vee P_{1,2} \vee P_{2,1}) \wedge [(\neg P_{1,2} \wedge \neg P_{2,1}) \vee B_{1,1}]$

4. Apply distributivity law ($\vee$ over $\wedge$) and flatten

$(\neg B_{1,1} \vee P_{1,2} \vee P_{2,1}) \wedge (\neg P_{1,2} \vee B_{1,1}) \wedge (\neg P_{2,1} \vee B_{1,1})$

26

26