Foundations of Machine Learning and Data Mining

Lecturer: Rainer Marrone, Prof. Dr. Ralf Möller
Lab classes: Rainer Marrone

Lecture: Thursdays (90 minutes)
Lab classes: Fridays (60 minutes)

Prerequisite:
Basic Mathematics

Recap

Probability is a rigorous formalism for uncertain knowledge

Joint probability distribution specifies probability of every atomic event

Queries can be answered by summing over atomic events

For nontrivial domains, we must find a way to reduce the joint size

Independence and conditional independence provide the tools

Slide: St. Russell

	toothache		¬toothache	
	pcatch	¬pcatch	pcatch	¬pcatch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

P(cavity|toothache) = P(cavity∧toothache)/P(toothache)

= (0.108+0.012)/(0.108+0.012+0.016+0.064) = 0.6

Interpretation: After observing toothache, the patient is no longer an "average" one, and the prior probabilities of cavity is no longer valid

P(Cavity|Toothache) is calculated by keeping the ratios of the probabilities of the 4 cases unchanged, and normalizing their sum to 1

Slide: Jean Claude Latombe

	toothache		¬toothache	
	pcatch	¬pcatch	pcatch	¬pcatch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

$P(Cavity|Toothache) = P(Cavity \wedge Toothache)/P(Toothache)$
$= (0.108+0.012)/(0.108+0.012+0.016+0.064) = 0.6$

$P(\neg Cavity|Toothache)=P(\neg Cavity \wedge Toothache)/P(Toothache)$
$= (0.016+0.064)/(0.108+0.012+0.016+0.064) = 0.4$

$$
\begin{aligned}
P(c|Toothache) &= \alpha\, P(c \wedge Toothache) \\
&= \alpha\, \Sigma_{pc}\, P(c \wedge Toothache \wedge pc) \\
&= \alpha\, [(0.108, 0.016) + (0.012, 0.064)] \\
&= \alpha\, (0.12, 0.08) = (0.6, 0.4)
\end{aligned}
$$

normalization constant

Slide: Jean Claude Latombe

Conditional Probability

- $P(A \wedge B) = P(A|B)\,P(B)$
  $= P(B|A)\,P(A)$
- $P(A \wedge B \wedge C) = P(A|B,C)\,P(B \wedge C)$
  $= P(A|B,C)\,P(B|C)\,P(C)$
- $P(Cavity) = \Sigma_t \Sigma_{pc}\, P(Cavity \wedge t \wedge pc)$
  $= \Sigma_t \Sigma_{pc}\, P(Cavity|t,pc)\,P(t \wedge pc)$
- $P(c) = \Sigma_t \Sigma_{pc}\, P(c \wedge t \wedge pc)$
  $= \Sigma_t \Sigma_{pc}\, P(c|t,pc) P(t \wedge pc)$

Slide: Jean Claude Latombe

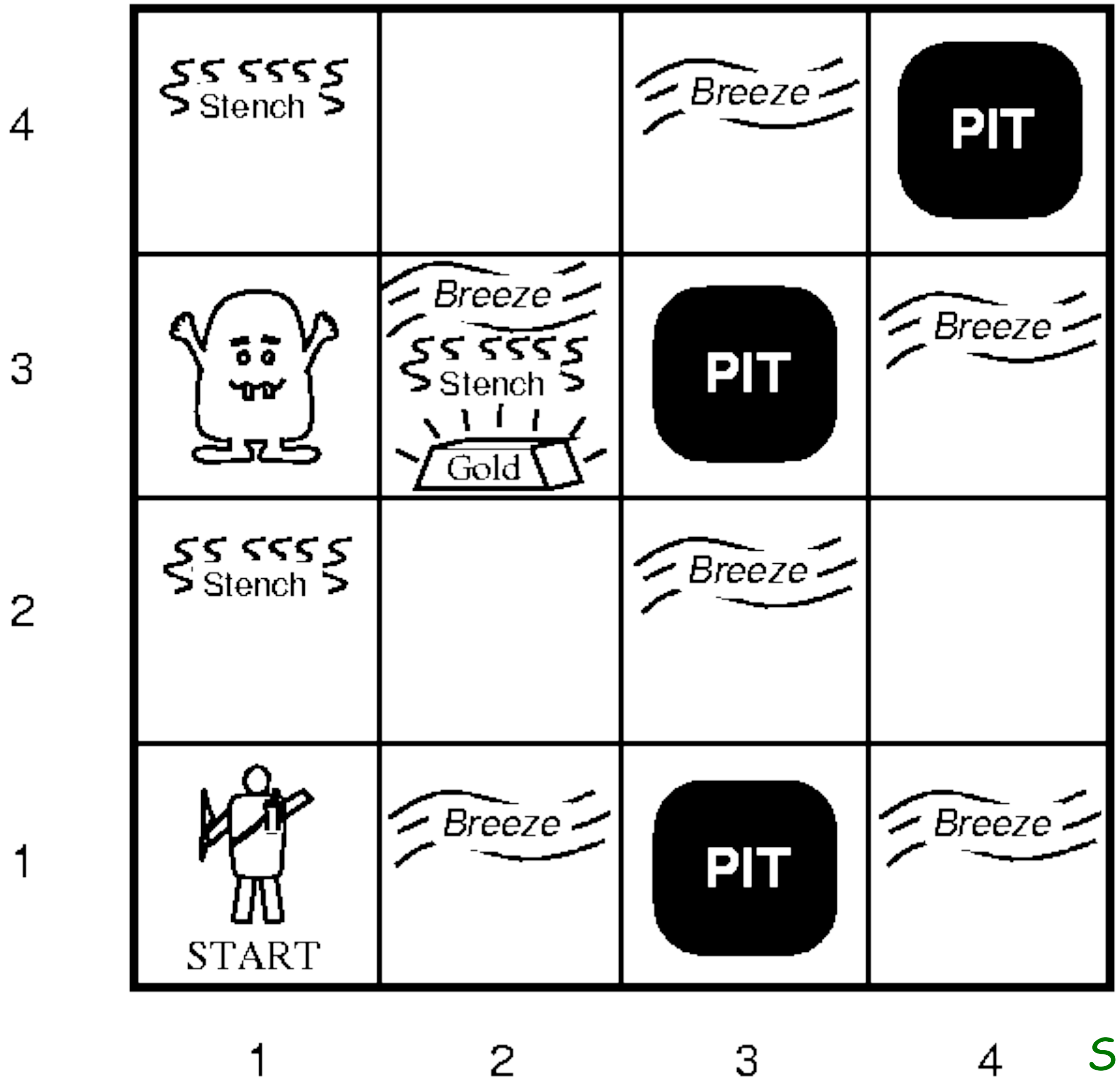
Independence

- Two random variables A and B are independent if
$$P(A \wedge B) = P(A)\,P(B)$$
hence if $P(A|B) = P(A)$

- Two random variables A and B are independent given C, if
$$P(A \wedge B|C) = P(A|C)\,P(B|C)$$
hence if $P(A|B,C) = P(A|C)$

Slide: Jean Claude Latombe

Wumpus World



Wumpus: Monster (death)

Pit: (death)

Stench: (near Wumpus)

Breeze: (near pit)

Glitter: (goal)

Hit: Agent moves against wall

Agent can move, turn, and shoot one arrow in the the direct it heads to

Agent has one arrow that kills the Wumpus (scream)

Slide: St. Russell

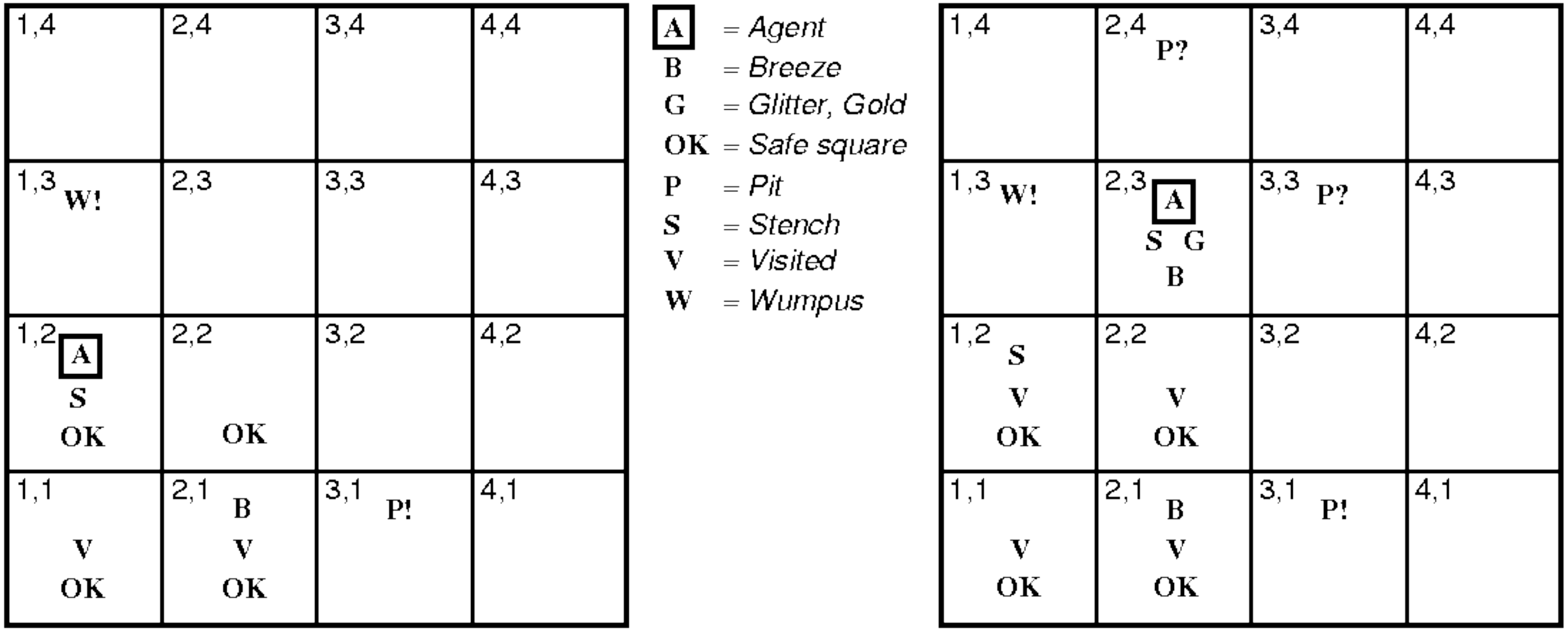
Wumpus World (2)

[1, 2] und [2, 1] sind sicher:

(a) (b)

Slide: St. Russell

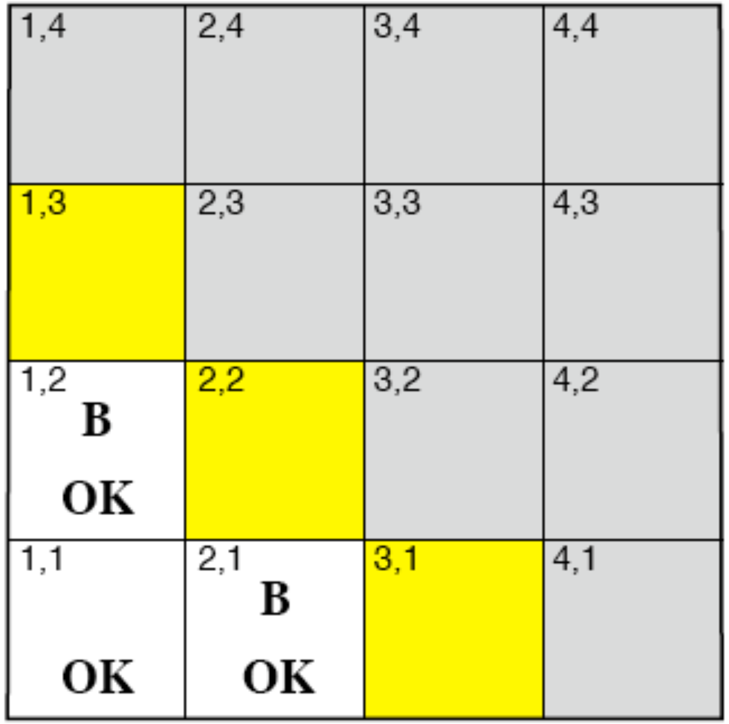
Wumpus Welt (3)



(a) (b)

Slide: St. Russell

Probabilistic Wumpus World



$P_{ij} = true$ iff $[i, j]$ contains a pit

$B_{ij} = true$ iff $[i, j]$ is breezy

Include only $B_{1,1}, B_{1,2}, B_{2,1}$ in the probability model

Slide: St. Russell

Specifying the probability model

The full joint distribution is $\mathbf{P}(P_{1,1}, \ldots, P_{4,4}, B_{1,1}, B_{1,2}, B_{2,1})$

Apply product rule: $\mathbf{P}(B_{1,1}, B_{1,2}, B_{2,1} \mid P_{1,1}, \ldots, P_{4,4})\mathbf{P}(P_{1,1}, \ldots, P_{4,4})$

(Do it this way to get $P(Effect|Cause)$.)

First term: 1 if pits are adjacent to breezes, 0 otherwise

Second term: pits are placed randomly, probability 0.2 per square:

$$\mathbf{P}(P_{1,1}, \ldots, P_{4,4}) = \Pi_{i,j=1,1}^{4,4}\mathbf{P}(P_{i,j}) = 0.2^3 \times 0.8^{16-3}$$

for 3 pits.

Slide: St. Russell

Observations and query

We know the following facts:

$b = \neg b_{1,1} \wedge b_{1,2} \wedge b_{2,1}$

$known = \neg p_{1,1} \wedge \neg p_{1,2} \wedge \neg p_{2,1}$

Query is $\mathbf{P}(P_{1,3}|known, b)$

Define $Unknown = P_{ij}$s other than $P_{1,3}$ and $Known$

For inference by enumeration, we have

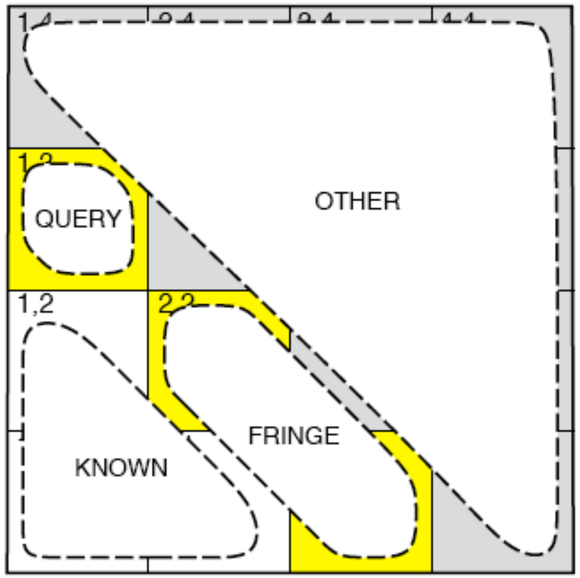
$$\mathbf{P}(P_{1,3}|known, b) = \alpha \Sigma_{unknown} \mathbf{P}(P_{1,3}, unknown, known, b)$$

Grows exponentially with number of squares!

Slide: St. Russell

Using conditional independence

Basic insight: observations are conditionally independent of other hidden squares given neighbouring hidden squares



Define $Unknown = Fringe \cup Other$
$\mathbf{P}(b|P_{1,3}, Known, Unknown) = \mathbf{P}(b|P_{1,3}, Known, Fringe)$

Manipulate query into a form where we can use this!

Slide: St. Russell

Using conditional independence (2)

$$\mathbf{P}(P_{1,3}|known, b) = \alpha \sum_{unknown} \mathbf{P}(P_{1,3}, unknown, known, b)$$

$$= \alpha \sum_{unknown} \mathbf{P}(b|P_{1,3}, known, unknown)\mathbf{P}(P_{1,3}, known, unknown)$$

$$= \alpha \sum_{fringe} \sum_{other} \mathbf{P}(b|known, P_{1,3}, fringe, other)\mathbf{P}(P_{1,3}, known, fringe, other)$$

$$= \alpha \sum_{fringe} \sum_{other} \mathbf{P}(b|known, P_{1,3}, fringe)\mathbf{P}(P_{1,3}, known, fringe, other)$$

$$= \alpha \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe) \sum_{other} \mathbf{P}(P_{1,3}, known, fringe, other)$$

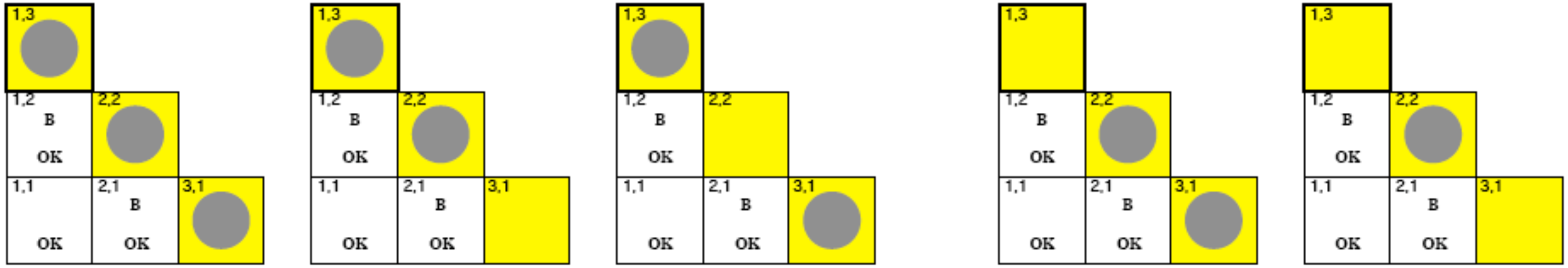
$$= \alpha \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe) \sum_{other} \mathbf{P}(P_{1,3})P(known)P(fringe)P(other)$$

$$= \alpha\, P(known)\mathbf{P}(P_{1,3}) \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe)P(fringe) \sum_{other} P(other)$$

$$= \alpha'\, \mathbf{P}(P_{1,3}) \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe)P(fringe)$$

Slide: St. Russell

Using conditional independence (3)



0.2 x 0.2 = 0.04 | 0.2 x 0.8 = 0.16 | 0.8 x 0.2 = 0.16 | 0.2 x 0.2 = 0.04 | 0.2 x 0.8 = 0.16

$$\mathbf{P}(P_{1,3}|known, b) = \alpha' \langle 0.2(0.04 + 0.16 + 0.16),\ 0.8(0.04 + 0.16) \rangle$$
$$\approx \langle 0.31, 0.69 \rangle$$

$$\mathbf{P}(P_{2,2}|known, b) \approx \langle 0.86, 0.14 \rangle$$

Slide: St. Russell