Reasoning and Decision-Making under Uncertainty

4. Session:
Beliefs, Evidence & Independence

Prof. Dr.-Ing. Stefan Kopp
Center of Excellence „Cognitive Interaction Technology"
AG Sociable Agents

CITEC

Sociable Agents

Recap: Degree of belief as probability

Degree of belief or probability of a *world* $Pr(\omega)$

Degree of belief or probability of a *fact (sentence)* $Pr(\alpha) := \sum_{\omega \models \alpha} Pr(\omega)$

State of belief or joint probability distribution

World	**Earthquake**	**Burglary**	**Alarm**	*Pr(.)*
w1	*true*	*true*	*true*	*.0190*
w2	*true*	*true*	*false*	*.0010*
w3	*true*	*false*	*true*	*.0560*
w4	*true*	*false*	*false*	*.0240*
w5	*false*	*true*	*true*	*.1620*
w6	*false*	*true*	*false*	*.0180*
w7	*false*	*false*	*true*	*.0072*
w8	*false*	*false*	*false*	*.7128*

$$\sum_{\omega_i} Pr(\omega_i) = 1$$

$$Pr(Earthquake) = .1$$
$$Pr(Burglary) = .2$$
$$Pr(Alarm) = .2442$$

Recap: Properties of degrees of beliefs

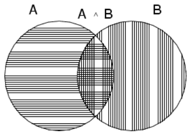
Properties

- bound $0 \leq Pr(\alpha) \leq 1 \quad \forall\alpha$
- baseline for inconsistent sentences $Pr(\alpha) = 0 \; \forall\alpha \; inconsistent$
- baseline for valid sentences $Pr(\alpha) = 1 \quad \forall\alpha \; valid$

Junctions of beliefs

- disjunction $Pr(\alpha \vee \beta) = Pr(\alpha) + Pr(\beta) - Pr(\alpha \wedge \beta)$
- conjunction $Pr(\alpha \wedge \beta) = \sum_{\omega \models \alpha, \omega \models \beta} Pr(\omega)$

$$Pr(\alpha \wedge \beta) = 0 \; if \; \alpha, \beta \; mutually \; exclusive$$



$$Pr(Earthquake \wedge Burglary) = Pr(\omega_1) + Pr(\omega_2) = .02$$

$$Pr(Earthquake \vee Burglary) = .1 + .2 - .02 = .28$$

Recap: Updating beliefs

new evidence β ➔ update state of beliefs such that

- worlds that contradict evidence get zero prob.
- worlds that had zero prob continue to have zero prob.
- worlds consistent with evidence and positive prob. maintain relative beliefs

conditioning of belief in a world ω :

$$Pr(\omega|\beta) := \begin{cases} 0 & \omega \models \neg\beta \\ \frac{Pr(\omega)}{Pr(\beta)} & \omega \models \beta \end{cases}$$

Bayesian conditioning of belief in event α:

$$Pr(\alpha|\beta) = \frac{Pr(\alpha \wedge \beta}{Pr(\beta)}$$

Belief dynamics under incoming evidence is a consequence of the state of beliefs one had

Chain rule & total probability

Repeated application of Bayes Conditioning gives chain rule

$$Pr(\alpha_1 \wedge \alpha_2 \wedge ... \wedge \alpha_n) = Pr(\alpha_1|\alpha_2 \wedge ... \wedge \alpha_n)Pr(\alpha_2|\alpha_3 \wedge ... \wedge \alpha_n)...Pr(\alpha_n)$$

If events β_i are mutually exclusive and exhaustive, we can apply case analysis or law of total probability to compute a belief in α :

$$Pr(\alpha) = \sum_i Pr(\alpha \wedge \beta_i) = \sum_i Pr(\alpha|\beta_i)Pr(\beta_i)$$

- compute belief in α by adding up beliefs in exclusive cases $\alpha \wedge \beta_i$ that cover the conditions under which α holds

„marginalizing over beta“
„Pr(a) marginal prob. of A“

Thomas Bayes (1702–1761)

Proposed a solution to the problem of "inverse probability"

- published posthumously by R. Price in Phil. Trans. of Royal Soc. Lond. (1763)

Bayes' theorem

- expresses posterior (after evidence *E* is observed) of a hypothesis *H* in terms of the priors of *H* and *E*, and the prob of *E* given *H*
- subjective interpretation implies that
  - strong beliefs are confirmed more by consistent evidence
  - evidence has a stronger effect if more unlikely before being observed

$$Pr(\alpha|\beta) = \frac{Pr(\beta|\alpha)P(\alpha)}{Pr(\beta)}$$

Bayes rule

Excercise:

A patient has been tested positive for a disease D, which one in every 1000 people has. The test T is not reliable (2% false positive rate and 5% false negative rate). What is our degree of belief *Pr(D|T)*?

$$Pr(D) = 1/1000$$

$$Pr(T|\neg D) = 2/100 \quad \Rightarrow \quad Pr(\neg T|\neg D) = 98/100$$

$$Pr(\neg T|D) = 5/100 \quad \Rightarrow \quad Pr(T|D) = 95/100$$

$$P(D|T) = \frac{95/100 \cdot 1/1000}{Pr(T)}$$

$$P(T) = Pr(T|D)Pr(D) + Pr(T|\neg D)Pr(\neg D)$$

$$Pr(D|T) = \frac{95}{2093} = 4.5\%$$

Soft & hard evidence

Often useful to distinguish two types of evidence

- hard evidence: information that some event has occurred
- soft evidence: unreliable hint that an event have may occurred
  - neighbor with hearing problem tells us he had heard the alarm trigger
  - can be interpreted in terms of noisy sensors

How to update in light of soft evidence? *Two methods*:

1. new state of beliefs *Pr'* = old beliefs + new evidence („all things considered") ➔ bayesian conditioning leads to Jeffrey's rule:

$$Pr'(\alpha) = qPr(\alpha|\beta) + (1-q)Pr(\alpha|\neg\beta) \text{ with } Pr'(\beta) = q$$

$$Pr'(\alpha) = \sum_i q_i Pr(\alpha|\beta_i) \text{ with } q_i \text{ exclusive and exhaustive}$$

Soft & hard evidence

2. use strength of evidence, independent of current beliefs („nothing else considered“)

- *Definition*: Odds of event: $O(\beta) := \frac{Pr(\beta)}{Pr(\neg\beta)}$
  - states how many times we believe more in β than in $\neg\beta$
- *Definition*: Bayes factor of the „strength“ of evidence: $k := \frac{O'(\beta)}{O(\beta)}$
  - relative change induced on odds of β
  - k=1: indicates neutral evidence
    k=2: indicates evidence strong enough to double the odds of β
    k➔∞: hard evidence conforming β, k➔0: hard evidence against β
- update according to evidence β with known Bayes factor *k*:

$$Pr'(\beta) = \frac{kPr(\beta)}{kPr(\beta)+Pr(\neg\beta)}$$

(from def. of *O*)

$$Pr'(\alpha) = \frac{kPr(\alpha\wedge\beta)+Pr(\alpha\wedge\neg\beta)}{kPr(\beta)+Pr(\neg\beta)}$$

(together with Jeffrey's rule)

Soft evidence

Example: Murder with three suspects, investigator Rich has the following state of belief:

- *O*(killer=david) = *Pr*(david)/*Pr*(not david) = 2

	Killer	*Pr(.)*
ω_1	david	2/3
ω_2	dick	1/6
ω_3	jane	1/6

new soft evidence is obtained that triples the odds of killer=david

- Bayes factor *k*=*O*'(killer=david)/*O*(killer=david) = 3

➔ new belief in David being the killer:

- *Pr*'(killer=david) = (3*2/3) / (3*2/3+1/3) = 6/7

simply the first statement (*k*; „nothing else considered“) can be used by other agents to update their belief according to β

Properties of beliefs

Key observation: Full joint distribution or state of belief is sufficient to model uncertain beliefs and update them in face of any kind of evidence

- determines prob for every event given *any combination* of evidence
- that is, enables all kinds of probabilistic inferences

Unfortunately, the joint distribution is exponential and therefore costly

- $O(d^n)$ with n random variables and domain size d

Idea: Exploit independencies in the world, i.e. assumptions that certain variables have nothing to do with each other, and learning about one doesn‘t change (degree of) belief in the other

- *„our most basic, robust, and commonly available knowledge about uncertain environments“*

Independence

A given state of beliefs finds a belief independent of another belief **iff**

$$Pr(\alpha|\beta) = Pr(\alpha) \text{ or } Pr(\beta) = 0$$

Equivalent definition (using product rule): $Pr(\alpha \wedge \beta) = Pr(\alpha) \cdot Pr(\beta)$

Examples & properties:

- initial state of beliefs, as defined above:
  - *Pr*(*Earthquake*)=.1 and *Pr*(*Earthquake* | *Burglary*)=.1
  - *Pr*(*Burglary*)=.2 and *Pr*(*Burglary* | *Earthquake*)=.2

  ➔ *Earthquake* and *Burglary* are independent

  ➔ knowing one doesn‘t change degree of belief in the other
- independence is always symmetrical, but different from mutual exclusiveness (of events)

Conditional Independence

<u>*Observation:*</u> **Independence is a dynamical notion**

- *Earthquake* and *Burglary* get dependent with evidence about *Alarm*
  - *Pr*(*Burglary* | *Alarm*)=.741 and *Pr*(*Burglary* | *Alarm*∧*Earthquake*)=.253

  ➔ *Earthquake* changes the belief in *Burglary* in presence of *Alarm*
- new evidence can make independent beliefs dependent, and vice versa!

<u>*Definition:*</u>

state of belief Pr finds α conditionally independent of β given event γ **iff**

$$Pr(\alpha|\beta \wedge \gamma) = Pr(\alpha|\gamma) \text{ or } Pr(\beta \wedge \gamma) = 0$$

- conditional independence is always symmetric

$$Pr(\alpha \wedge \beta|\gamma) = Pr(\alpha|\gamma)Pr(\beta|\gamma) \text{ or } Pr(\gamma) = 0$$

Conditional Independence

<u>*Example*</u>:

Given two noisy, unreliable sensors

Temp	*sensor1*	*sensor2*	*Pr(.)*
normal	*normal*	*normal*	*.576*
normal	*normal*	*extreme*	*.144*
normal	*extreme*	*normal*	*.064*
normal	*extreme*	*extreme*	*.016*
extreme	*normal*	*normal*	*.008*
extreme	*normal*	*extreme*	*.032*
extreme	*extreme*	*normal*	*.032*
extreme	*extreme*	*extreme*	*.128*

Initial beliefs

- *Pr(Temp=normal)=.80*
- *Pr(Sensor1=normal)=.76*
- *Pr(Sensor2=normal)=.68*

After checking sensor1 and finding its reading is *normal*

- *Pr(Sensor2=normal | Sensor1=normal) ~ .768* ➔ *Sensor1* and *Sensor* 2 dependent

After observing that temperature is *normal*

- *Pr(Sensor2=normal |Temp=normal)* = *.80*
- *Pr(Sensor2=normal | Temp=normal, Sensor1=normal)* = *.80* ➔ cond. independent

Conditional Independence

Is a special case of mutual information, which quantifies the impact of observing one variable on the uncertainty in another:

$$MI(X;Y) := \sum_{x,y} Pr(x,y) log_2 \frac{Pr(x,y}{Pr(x)Pr(y)}$$

- non-negative
- zero iff *X* and *Y* are independent

Relation to entropy:

$$MI(X;Y) := ENT(X) - ENT(X|Y) = ENT(Y) - ENT(Y|X)$$

- with conditional entropy: $ENT(X|Y) := \sum_y Pr(y) log_2 ENT(X|y)$
  $ENT(X|y) := -\sum_x Pr(x|y) log_s Pr(x|y)$
- Note: $ENT(X|Y) \leq ENT(X)$

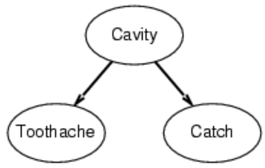
Independence of (sets of) variables

Notation:
independence between sets of variables **X,Y, Z** in a belief state *Pr*
denoted as $I_{Pr}(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$

Example:

- **X**={A,B}, **Y**={D,E}, **Z**={C}
- $I_{Pr}(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ denotes 4x2x4=32 different independent statements:
  - A∧B indep. of C given D∧E
  - A∧¬B indep. of C given D∧E
  - ..
  - ..
  - ¬A∧¬B indep. of ¬C given ¬D∧¬E

Why is independence so neat?



It allows to decompose a joint distribution!

- *Pr(Cavity,Catch,Toothache)* ➔ 2^3=8 worlds needed
  = *Pr(Tootha.,Catch|Cavity) Pr(Cavity)* (Product rule)
  = *Pr(Tootha.|Cavity) Pr(Catch|Cavity) Pr(Cavity)* (cond. ind. of *Toothache* & *Catch* given *Cavity*) ➔ 2+2+1=5 worlds needed

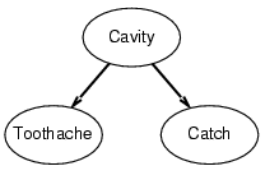
Common pattern:

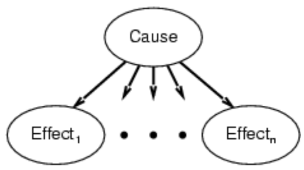
If a cause directly implies multiple effects, all of which are conditionally independent given the cause, then:

$$Pr(Cause, E_1, ..., E_n) = Pr(Cause) \prod_i Pr(E_i|Cause)$$

- the cause sufficiently „explains" each effect, knowing about other effects doesn't change the belief in it anymore
- Naive Bayes model (also called Bayesian classifier):
  Bayes rule + presumed independence where there is no

Bayesian networks

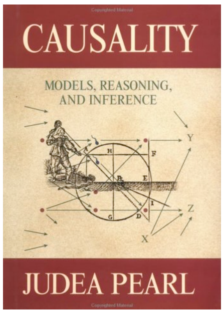




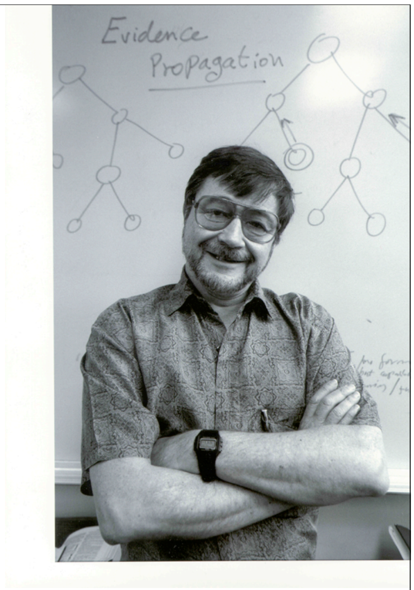
Definition: A Bayesian network for variables Z is a pair (G,Θ) with

- Structure G: a directed acyclic graph with
  - a set of nodes, one per random variable
  - a set of edges representing *direct causal influence* between variables
- Parametrization Θ: a conditional probability table (CPT) for each variable (node) given its parents:
  - *Pr(X_i | Parents(X_i))* or *Pr(X_i)* if there are no parents
  - parameterizes the independence structure

(1988)

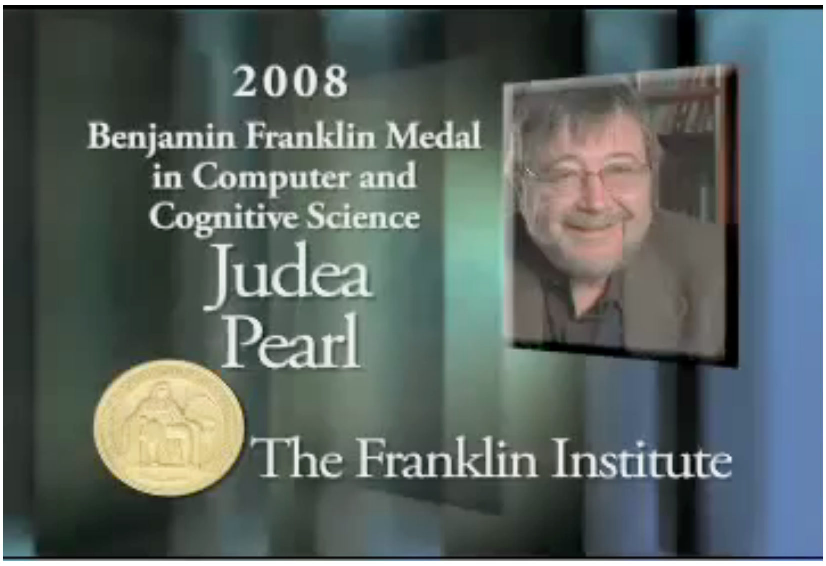


(2000)



Judea Pearl coined the term Bayesian networks to emphasize:

- the subjective nature of the input information
- the reliance on Bayes's conditioning as the basis for updating beliefs
- the distinction between causal and evidential modes of reasoning



ACM Awards Judea Pearl the Turing Award for Work On Artificial Intelligence

Mar 15, 2012 11:54 AM EST | 2 Comments

By Michael J. Miller

Summary

Belief updating using probability theory

- chain rule, law of total probability
- Bayes‘ rule

Update under evidence

- Hard evidence: conditioning
- Soft evidence: Jeffrey‘s rule, odds, Bayes factor

Independence and mutual information

- symmetrical, dynamic
- unconditional and conditional independence
- allows to decompose joint probability distributions
  (simplest form: naive Bayes classifier)