Information

Torsten Enßlin 22.7.24

What is information?

Latin root "informare" = to give shape

What is shaped? Knowledge!

What is Knowledge? Internal representation in a real or virtual agent's mind of properties of the world!

s = state $\in S$

S = set of possible states

agent = human, animal, cell, virtual entity that is imagined to have created the decisions made by evolution

Knowledge s_A allows to plan action a_A according to utility function $u_A(s, a)$.

E.g. $a_A = \operatorname{argmax}_a u_A(s_A, a)$

Problem: s is not known, s_A could be wrong.

Solution: quantification of certainty that $s \in S$ is the case

Requirements of certainty quantification:

- logic should be limiting case in case of certain states
- sequence order in which pieces of information are included in reasoning is irrelevant
- calibration: known frequencies of events should be their certainty

Cox's theorem

⇒ Probabilities: $P: \mathcal{S} \to [0,1]$

$\mathcal{S}' = 2^S = \{0,1\}^S$ = power set of S = set of all subsets

event

$P(s \in u \mid s \in \mathcal{S}' \supset u) \in [0,1]$, $P(s \in \mathcal{S}' \mid s \in \mathcal{S}') = 1$, $P(s \in \{\} \mid s \in \mathcal{S}') = 0$

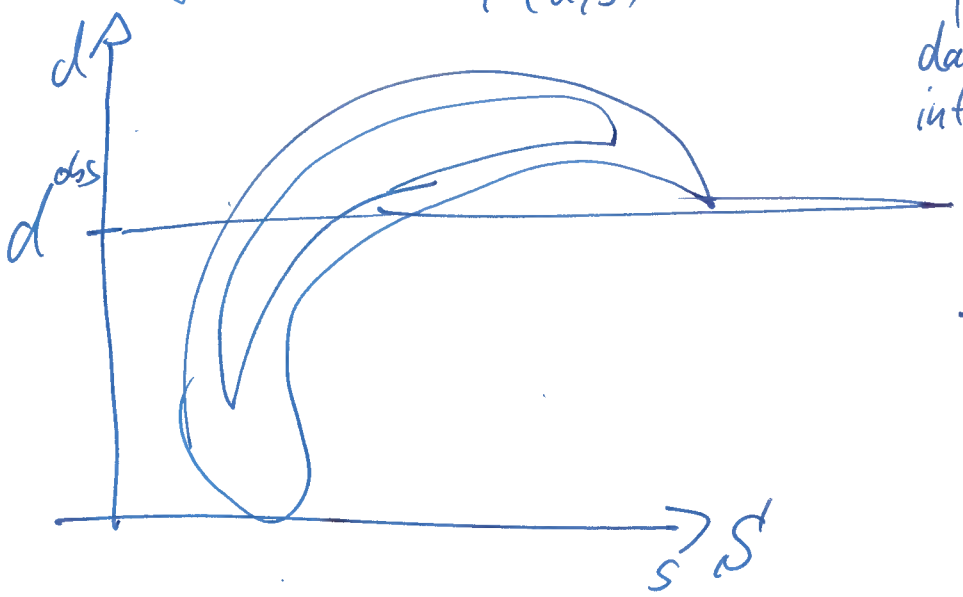
$P(\text{statement} \mid \text{condition})$

calibration:

$P(\text{event } e \text{ happens} \mid \underbrace{e \text{ has frequency } f}) = f$

$f = F(e) = \frac{\#\, e \text{ happens}}{\#\, \text{anything happens}}$

Updating



$P(s) \to P(s|d)$

↑ data as information source

$$P(s|d=d^{obs}) = \frac{P(d,s)}{P(d)}\Big|_{d=d^{obs}}$$

Bayes

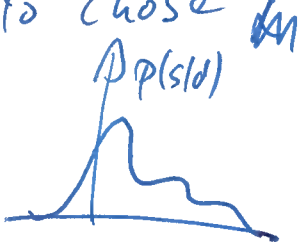
posterior $P(s|d)$ might be complex

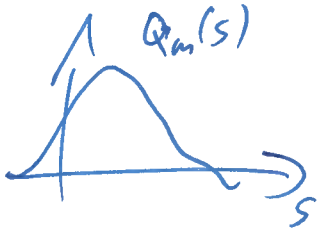
simplified, approx, or compressed representation

$$Q_{dm}(s) \simeq P(s|d)$$

↑ some parameters, compressed data

How to chose dm?





Alice

Bob (e.g. future version of Alice and m is then memory)

some loss or scoring function $D_s(d,m) \equiv D_s(P\|Q)$

Principles for scoring rule:

1) be proper if $Q_m(s) = P(s|d)$ is possible do so
2) ~~be optimal~~ maximal: $Q_m(s)$ for s that happens

$\Rightarrow$ ~~Maximize~~ $\mathcal{L}(Q_m(s))\big\rangle_{(s|d)}$

Alice knowledge

3) ~~be~~ differential
4) be calibrated: $D_s(P\|P) = 0$

$$\Rightarrow D_s(d,m) = \int ds\, P(s|d) \ln \frac{P(s|d)}{Q_m(s)}$$

(1) relative Entropy = # nits lost by $P \to Q$

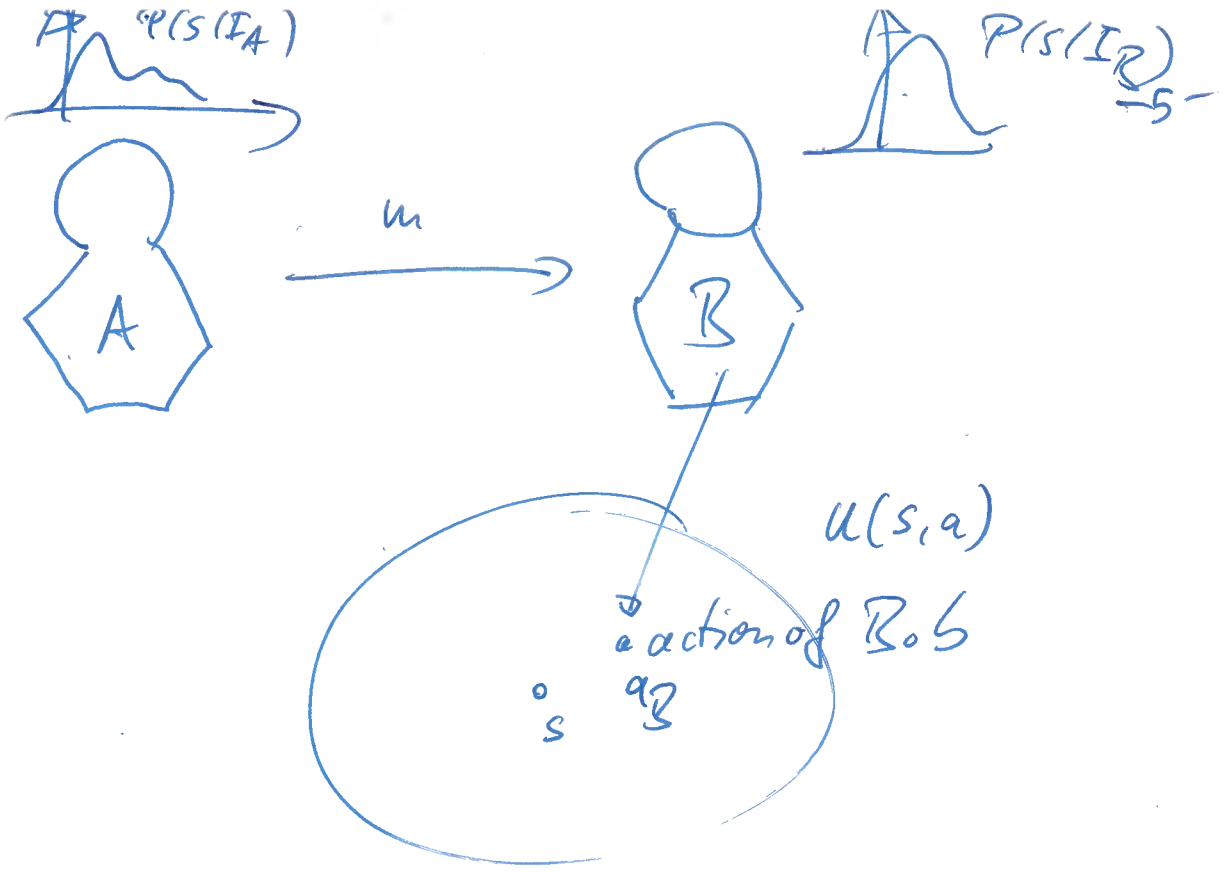
$Q_m(s) = \text{const}$ $D_s \to$ Shannon entropy

other scoring rules are possible:

relative entropy does not reflect relevance of information measured

solution: attention $\mathcal{A}(s|I) = \frac{w(s)P(s|I)}{\int ds\, w(s) P(s|I)}$

(2) attention entropy $D_s(\mathcal{A}\|\mathcal{A}'')$ is proper, differentiable, calibrated and attention optimal.



If Alice knows $u(s,a)$

maximize $\langle u(s, a_B) \rangle_{(s|I_A)}$

where $a_B = \arg\max_a \langle u(s,a) \rangle_{(s|I_B)}$

$\Rightarrow$ Alice chooses message (or $P(s|I_B)$) that let's Bob do what she think is best

$\Rightarrow$ ~~measure~~ scoring rule is not proper

If Alice does not know $u(s,a)$, but just that has maxima with curvature $w(s)$

$$u(s,a) = u_0(s) + \frac{w(s)}{2}(a - a_*(s))^2$$

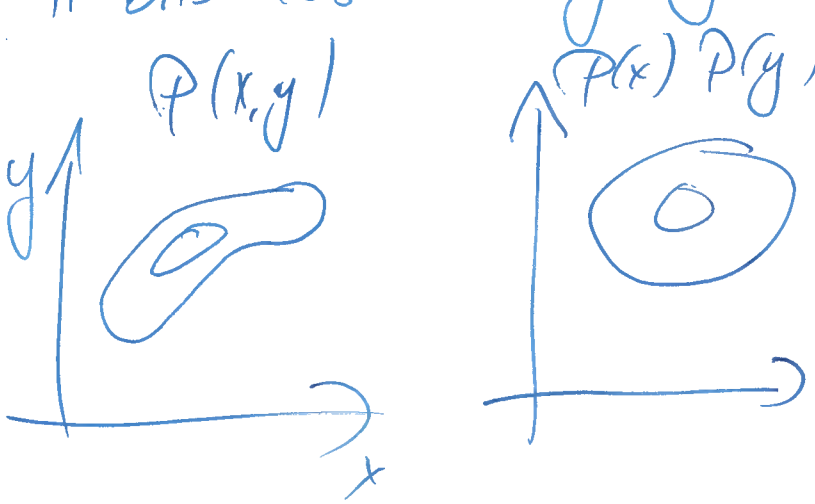
$\Rightarrow$ $\int ds \left[P(s|I_A) - P(s|I_B) \right]^2$ should be weighted (Brier-) minimized score

mutual information of x and y

$$MI = D_{(x,y)}\left(P(x,y) \,\|\, P(x)P(y)\right)$$

$$= \int ds\, P(x,y) \ln \frac{P(x,y)}{P(x)P(y)}$$

⇒ ~~how much info~~ # bits lost in forgetting relation of x to y



surprisal $H(s|I) = -\ln P(s|I)$ information potential/ Hamiltonian

"additivity of information"

product rule $P(x,y) = P(x|y)\,P(y)$

$$\rightarrow H(x,y) = H(x|y) + H(y)$$

$$P(s|d) = \frac{P(s,d)}{P(d)} = \frac{e^{-H(s,d)}}{Z(a)}$$

information theory = statistical mechanics

Summary

Information means change of knowledge

Knowledge is described by probabilities

data = something we know for sure

can change knowledge

prior $P(s) \longrightarrow P(s|d)$ posteria

relative entropy quantifies amount of information lost in imperfect communication/memorization/processing.

entropy is allways relative to some reference probability distribution

(specify space S and reference $Q(s)$ when mentioning an entropy!)

entropy is a scoring rule [expressing closedness of knowledge states]

out of a rich zoo of scoring rules

[for other purposes].

entropy is a negative **amount of information**

Other usage of word "information": mutual information, surprisal = information potential function, data, ...