Logic of Learning

Radboud University Nijmegen
KNAW-webinar, Logic of AI, 22 April 2021

Bart Jacobs
bart@cs.ru.nl

The end of Theory (Wired, 2008)

Chris Anderson published a provocative, influential article *The end of Theory. The Data Deluge makes the Scientific Method Obsolete.*

- . . . faced with massive data, this approach to science — hypothesize, model, test — is becoming obsolete.
- There is now a better way. Petabytes allow us to say: "Correlation is enough." We can stop looking for models. We can analyze the data without hypotheses about what it might show.

Thus:
- no causation but correlation
- no Boolean logic, but what logic instead?

See: www.wired.com/2008/06/pb-theory

Logic of Learning

Correlation and causation

Example

Studies show that there is a strong correlation between:
- occurrence of lung cancer
- presence of ash trays in homes

Do ash trays cause cancer?

No, of course not!

Smoking is a "confounding factor". It's the common cause for both lung cancer and the presence of ash trays.

Anderson's claims that correlations suffice are too simplistic

See also: Pearl & Mackenzie, The Book of Why, 2019

Challenges in probabilistic logic (from Pearl'89)

> *To those trained in traditional logics, symbolic reasoning is the standard, and nonmonotonicity a novelty. To students of probability, on the other hand, it is symbolic reasoning that is novel, not nonmonotonicity. Dealing with new facts that cause probabilities to change abruptly from very high values to very low values is a commonplace phenomenon in almost every probabilistic exercise and, naturally, has attracted special attention among probabilists. The new challenge for probabilists is to find ways of abstracting out the numerical character of high and low probabilities, and cast them in linguistic terms that reflect the natural process of accepting and retracting beliefs.*

Embarrassingly, there is still **no** probabilistic logic for symbolic reasoning.

Probabilistic reasoning and updating (belief revision)

Example

I may think that scientists are civilised people. But then I attend a conference dinner that ends in a fist fight.

I will update my judgement.

- This is difficult in traditional, monotonic logic, where adding more information can not make true statements false.
- We need to switch from truth/falsity of statement, to likelihood
  - not two-element set $\{0, 1\}$ but interval $[0, 1]$
  - not sharp but fuzzy (soft) statements

  The likelihood that scientits are civilised is decreased, by the events at the conference dinner, through updating (belief revision).

Naive picture of learning

Alternative: predictive coding theory (Carl Friston et al)

- The human mind is constantly active in making predictions
- These predictions are compared with what actually happens
- Mismatches (prediction errors) lead to updates in the brain

"The human brain is a Bayesian prediction & correction engine"

My own (logical) interests/work

- There are two update rules, by Judea Pearl (1936) and by Richard Jeffrey (1926 – 2002)
  - What are the differences?
  - When to use which rule?
- The topic is mathematically non-trivial
  - also because I like category-theoretical methods ☺
  - details are skipped here
  - one point: difference arise with use of fuzzy statements
- Intriguing question: does the human mind use Pearl's or Jeffrey's rule — within predictive coding theory
  - cognition theory may provide an answer

See: BJ, *The Mathematics of Changing one's Mind, via Jeffrey's or via Pearl's update rule*, Journ. of AI Research, 2019

Burglary-alarm example

Consider the following probability table:

	burglary	no burglary
alarm	$1/200$	$7/500$
no alarm	$1/1000$	$98/100$

Somebody reports an alarm, with 80% certainty — e.g. because of deafness. What is the burglary probability?

- **According to Pearl:** $\frac{3}{151} \approx 2\%$
- **According to Jeffrey:** $\frac{19639}{93195} \approx 21\%$

This difference makes many experts uncomfortable.

Calculations (for aficionados)

Pearl: multiply the evidence probabilities into the table:

	burglary	no burglary
alarm	$8/10 \cdot 1/200$	$8/10 \cdot 7/500$
no alarm	$2/10 \cdot 1/1000$	$2/10 \cdot 98/100$

Then normalise and marginalise (or the other way around):

$$\frac{u}{u+v} = \frac{3}{151} \quad \text{for} \quad \begin{cases} u = 8/10 \cdot 1/200 + 2/10 \cdot 1/1000 \\ v = 8/10 \cdot 7/500 + 2/10 \cdot 98/100 \end{cases}$$

Jeffrey: Use $8/10 \cdot P(b \mid a) + 2/10 \cdot P(b \mid \text{not } a) = \frac{19639}{93195}$ where:

$$P(b \mid a) = \frac{1/200}{1/200 + 7/500} \qquad P(b \mid \text{not } a) = \frac{1/1000}{1/1000 + 98/100}$$

Comparison table about updating (with informal descriptions)

	Pearl's rule	Jeffrey's rule
effect	increase of what's right	decrease of what's wrong
you learn nothing from	uniformity (no differences)	what you already know (predict)
successive updates commute?	yes	no

One million dollar question

- Does the human mind use Pearl's or Jeffrey's rule?
- My bet is on Jeffrey ...
- Since the human mind is very sensitive to the order of updating (priming)

My favourite example: consider the impact of the following two sentences, in different orders.

Alice is pregnant Bob visits Alice.

Versus:

Bob visits Alice Alice is pregnant.

Conclusions

- Learning from data is becoming increasingly important
  - including subsequent classification and decision making
  - applied in many businesses and (public) organisations
- We hardly know what is happening
  - urgent need for a (symbolic) probabilistic logic — that includes updating
  - not only in AI, but also in cognition theory
  - when to apply Pearl or Jeffrey is poorly understood: interesting case study.

Thanks for listening; I hope you learned a bit.