Is there any Need to Mention Induction? CogSci 2010 Poster

Chris Thornton
Informatics
University of Sussex
Brighton
BN1 9QH
UK
c.thornton@sussex.ac.uk

July 12, 2011

Motivation

- Use of seen data to predict unseen data (i.e., *induction*) seems a critical part of cognition at every level.
- But there seems no way we can specify a principled mechanism for it.
- Hume notes all such specifications seem to become circular (end up 'going in a circle, and taking that for granted, which is the very point in question')
- NFL people point out that if you did specify a principled mechanism it would need to work in all domains, which dooms it to the performance of random guessing overall.
- Machine learning cheats by fixing assumptions on the unseen data (the IID assumption).

Popper revisited

- Popper gets around the problem by suggesting inductive effects can emerge implicitly from a mechanism that is principled but *not* inductive.
- Unfortunately, his suggested mechanism is hypothesis-falsification, which is impractical and unrealistic.

But his argument can be reconstructed using information theory.

(1) Establish a way of measuring the informational efficiency of representational constructs.

(2) Show that adoption of relatively efficient constructs yields inductive effects implicitly.

(3) A general disposition towards efficiency is then all we need. It's fully principled, in the sense of being disposed towards efficient use of resources. But it is more practical and realistic than falsification.

Efficient reconstruction of symbolic data

The amount of information expressed by symbolic data D.

$$I(D) = |D| \log n$$

Where two or more constructs have the same structure, the combination forms a ‘union’.
Information loss of union x.

$$H(x) = \sum_i \log |x_i|$$

D' is then used to label a *reconstruction* of D in which some constructs are replaced with symbols representing unions.

Reconstruction properties

Total information loss from a reconstruction:

$$H(D') = \sum_i H(D'_i)$$

Total symbol cost:

$$c(D') = |D'| + \sum_{x \in D'} |x|$$

Informational efficiency (mean content of symbols):

$$\bar{I}(D') = \frac{I(D) - H(D')}{c(D')}$$

White swans example

Data

large	white	flying	swan
large	white	swimming	swan
small	white	flying	swan
medium	white	swimming	swan
small	white	swimming	swan

Efficient reconstruction

-2.0	\$0 = small/large white flying/swimming swan
-2.0	\$0
-2.0	\$0
	medium white swimming swan
-2.0	\$0
-8.0	(40.0-8.0)/12 = 2.67 bits per symbol

More efficient reconstruction

-2.58	\$1 = medium/large/small white flying/swimming swan
-2.58	\$1
-2.58	\$1
-2.58	\$1
-2.58	\$1
-12.9	(40.0-12.92)/9 = 3.01 bits per symbol

Note effects of implicit induction.

Black ravens example

medium	black	flying	raven
large	white	flying	swan
small	white	flying	swan
medium	black	perching	raven
small	white	perching	swan
large	black	flying	raven
large	white	perching	swan
medium	white	perching	swan
small	black	flying	raven

Efficient reconstruction

-2.58	\$2 = medium/small/large black perching/flying raven
-2.58	\$3 = medium/large/small white perching/flying swan
-2.58	\$3
-2.58	\$2
-2.58	\$3
-2.58	\$2
-2.58	\$3
-2.58	\$3
-2.58	\$2
-23.26	(72.0-23.26)/17 = 2.87 bits per symbol

Conclusions

- This approach ‘solves’ the philosophical problem of induction to the same extent (and in the same way) falsification does. But more plausible assumptions of functionality are made.
- Hume’s circularlity is avoided, provided we take a disposition towards efficiency to be given (i.e., not something that is itself inductively acquired).
- NFL argument doesn’t apply because there no claim is made for universal induction. Efficiency-based induction is essentially a form of compression. As such it is feasible only with non-random data.
- As Popper argued, there’s no need ‘even to mention induction’ in showing the principle on which it operates.