Causal Inference reading group - summaries (ongoing)

Francisco Simoes

May 2021

Here I leave summaries of some of the papers that are analysed in the reading group of the causal inference lab from the University of Amsterdam's Institute of Logic, Language and Computation, which I joined in May 2021. I will update this file every couple of weeks.

1 A20210503 "Bayesian modeling of human concept learning" by Joshua Tenenbaum (1999) [1]

1.1 Context

Humans can successfully generalise from a small number of positive examples of a concept. Negative examples are useful but not necessary. Most machine learning algorithms are in stark contrast with this. It would be interesting to better model human concept learning.

1.2 Aim

Develop a theory for concept learning from few positive samples that performs well on the well-known task of learning axis-parallel rectangles in $\mathbb{R}^m$. Each rectangle is to be thought of as modelling a concept, and the learning problem is to infer what points belong to the rectangle, given a small set of points within the rectangle.

1.3 Contributions

- They develop a novel theoretical analysis of the rectangle learning problem based of Bayesian inference.
- Using experimental data, they show that this new approach is the best model of human performance on the rectangle problem, when compared with a selection of other models.

1.4 The method

To learn from positive samples only, they try to construct a model treating a new concept C as an unknown subset of "objects" (points of $\mathbb{R}^m$) and try to estimate C by finding a "good" hypothesis in the hypothesis space H of all possible concepts.

For the rectangle task, H is the set of all rectangles in the plane. Naturally, there are infinitely many rectangles consistent with any dataset of positive samples. But if we *assume* that the samples are generated by random sampling, then we can use Bayesian inference to say what hypotheses are more likely. This assumption dictates the form of the likelihood function $p(X \mid h)$. This gives rise to the *size principle*: from all the rectangles (hypotheses) consistent with the data, the smallest tend to be the most likely ones, and the likelihood increases exponentially with the number of positive samples.

Note: The size principle is not difficult to intuit: for example, if you have 6 positive samples (points that you know lie inside the rectangle C you want to estimate) on a huge 1 kilometre by 2 kilometres rectangular map and all of them lie within a few centimetres from each other, you can certainly say that a the entire map is a plausible C – after all, all the positive samples trivially lie inside C – but if the samples were randomly samples, this hypothesis becomes extremely unlikely. After all, it is intuitively very unlikely that all samples ended up so close together by pure chance if they were being sampled at random from within C and the rectangle C was the whole map.

The second ingredient of this method is a prior probability density function. They explore two priors: the trivial uninformative prior and the expected-size prior. The latter encodes the belief that there is an expected rectangle size for each dimension of the rectangle, thus having one parameter σ representing the expected scale of the rectangle.

Given a likelihood function and a prior density function one can compute the probability $p(y \in C \mid X)$ of a certain new object (point in $\mathbb{R}^m$) y belongs to the concept (rectangle) C, given the observations (set of positive samples) X.

1.5 Comparison with other models when modelling humans

This method was compared with the MIN method (which just selects the smallest rectangle containing all positive samples) and the weak Bayes method (which assumes random sampling on the entire $\mathbb{R}^m$, and thus does not give rise to the size principle). More details about these methods and the experiment that was conducted to compare the models are explained in the paper. For the purposes of this summary, it suffices to say that the bayesian approach proposed in the paper is able to model the human guesses by using an expected-size prior with $\sigma = 5$.

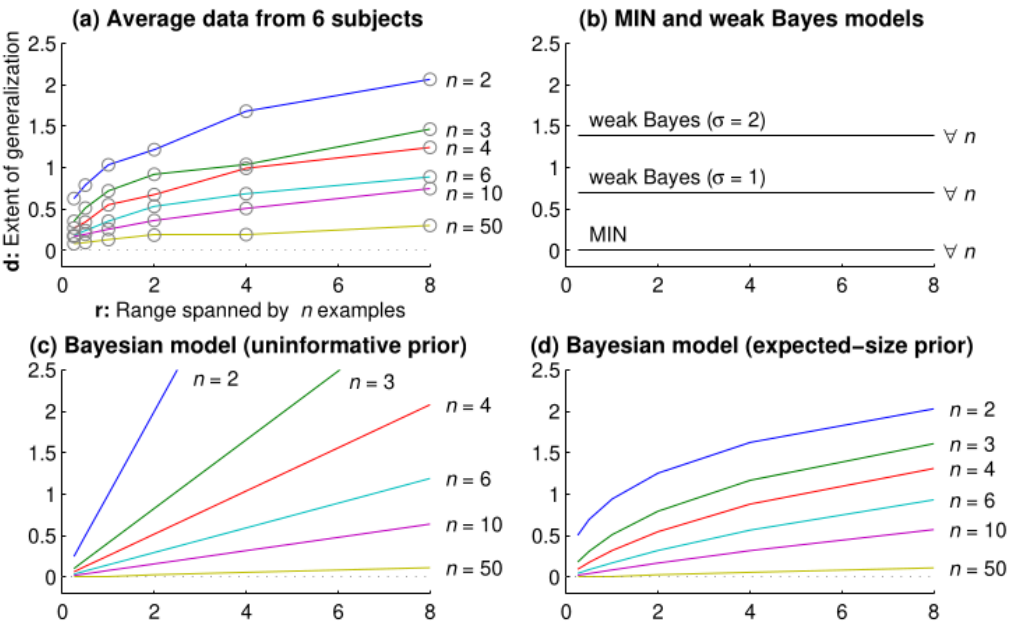

Figure 1: Comparison of human predictions and predictions from various models for the rectangle problem. Taken from [1].

1.6 A critical note

- The expected-size prior is introduced ad-hoc: there is no real reason to introduce it except that with it the model fits the data. This is not a critical issue since the point of the paper remains true: this model is able to mimic human performance on the rectangle task.
- Sample size from the experiment is very small.

References

[1] Joshua B Tenenbaum. Bayesian modeling of human concept learning. *Advances in neural information processing systems*, pages 59–68, 1999.