Inductive Logic 2015

UC Irvine

Analogical Predictions by Proximity among Predicates

Or: Inductive Logic on Attribute Space

⋆

Jan-Willem Romeijn

University of Groningen

Plan of this talk

Analogical predictions will be approached in two different ways.

- We review and reinterpret some older material involving similarities among predicates.
- The same mathematical tools are then applied on a refined observation space.

The move to a refined observation space is inspired by Marta Sznajder's project.

Eternal recurrence

The use of particular mathematical structures invokes the saying: “when you have a hammer, every problem looks like a nail”.

The reference to Nietzsche is entirely frivolous, yet fitting: in this talk we let go of fixed predicate categories.

Contents

1 Analogical predictions

Analogical predictions are predictions based on similar cases. Consider sampling pieces of fruit from a bag.

We categorize items i according to their fruit kind, $\{A_i, B_i, C_i, D_i\}$, and according to being round, R_i or $\bar{R}_i$, and having a stone, S_i or $\bar{S}_i$. Dates complete the picture but they are not needed in this talk.

Analogical predictions

Two different forms of analogy can be distinguished.

- Say we observe all round fruits to have a stone, and all non-round ones to not have one. Our next item is round. By analogy over items, we predict it will have a stone.
- Say we consider apples and cherries similar. We pick some cherries from the bag and so expect more cherries. By analogy over predicates, we also expect some apples.

Similarity by shared predicates

We can relate the two kinds of analogy by appealing to similarity or, more precisely, to their Hamming distance in terms of underlying predicates.

- Items i and i' are similar because they share a component predicate, namely R_i and $R_{i'}$.
- Predicates A and C are similar because they share a component predicate, namely R.

In what follows we focus mostly on the similarity of predicates.

Analogy among items

Carnap's predictions rules have the analogy among items built in. Say that

$$E_n = A_1 \cap B_2 \cap \ldots \cap A_{n-1} \cap B_n \quad \text{where} \quad A_i = R_i \cup \bar{S}_i \quad \text{and} \quad B_i = \bar{R}_i \cup \bar{S}_i.$$

Let P_C be a member of Carnap's continuum of inductive methods. Then we have

$$P_C\left(\bar{S}_{n+1} | R_{n+1} \cap E_n\right) \; > \; P_C\left(\bar{S}_{n+1} \cap R_{n+1} | E_n\right).$$

Analogy among predicates

There are many models of analogical predictions based on predicate similarity. In this talk we explore the idea of proximity among predicates, for example via shared underlying predicates. Note that

$$P(A_{n+1}|E_n) = P(R_{n+1}|E_n) \times P(\bar{S}_{n+1}|R_{n+1} \cap E_n).$$

We can separately specify the predictions of R_{n+1}, and of $\bar{S}_{n+1}$ conditional on R_{n+1}, as Carnapian rules with their own parameters.

Analogy among predicates

Stipulating that n_X denotes the number of occurrences of X in E_n, and assuming equal probability for all fruits at the outset, we can write

$$P(A_{n+1}|E_n) \;=\; \frac{n_R + \lambda\gamma_R}{n+\lambda} \times \frac{n_{\bar{S}|R} + \lambda_R\gamma_{\bar{S}|R}}{n_R + \lambda_R},$$

Seeing that $n_A = n_{\bar{S}|R}$ and choosing $\lambda_R = \lambda\gamma_R$ and $\gamma_A = \gamma_R\gamma_{\bar{S}|R}$, this reduces to the Carnapian rule P_C:

$$P_C(A_{n+1}|E_n) \;=\; \frac{n_A + \gamma_A\lambda}{n+\lambda}.$$

Differentiating learning rates

We can introduce an analogy between apples and cherries by choosing $\lambda_R > \lambda\gamma_R$:

$$P(A_{n+2}|C_{n+1} \cap E_n) > P(A_{n+2}|B_{n+1} \cap E_n).$$

Intuitively, finding a cherry (C_{n+1}) will make being round (R_{n+2}) more probable, and it makes having a stone (S_{n+2}) conditional on being round more probable. But by the differing learning rates, the former effect is far more pronounced.

2 Bayesian statistical underpinning

Prediction rules can also be derived from a Bayesian statistical treatment.

We consider hypotheses pertaining to possible proportions among apples, bananas and cherries in the fridge.

Bayesian statistical underpinning

The hypotheses H_θ determine the probability for drawing a fruit kind from the fridge according to:

$$P(X_{n+1}|H_\theta \cap E_n) = \theta_X,$$

where $X \in \{A, B, C\}$. We define a prior probability density $P(H_\theta)$, we use Bayesian conditioning to form the posterior, and we derive a prediction rule $P(A_{n+1}|E_n)$ from it:

$$P(A_{n+1}|E_n) = \int_{\theta\in\Theta} P(H_\theta|E_n)P(A_{n+1}|H_\theta \cap E_n)\, d\theta.$$

Bayesian statistical underpinning

The Carnapian predictions follow from this Bayesian procedure if we assume a Dirichlet prior:

$$P(H_\theta) \propto \theta_A^{\lambda\gamma_A - 1} \times \theta_B^{\lambda\gamma_B - 1} \times \theta_C^{\lambda\gamma_C - 1},$$

which we denote by $\mathrm{Dir}(\lambda\gamma_A, \lambda\gamma_B, \lambda\gamma_C)$. Upon observing a sequence E_n, we construct a posterior according to:

$$P(H_\theta) \propto \prod_{X \in \{A,B,C\}} \theta_X^{n_X + \gamma_X\lambda - 1}.$$

The exponents in the Dirichlet prior directly the Carnapian predictions.

Transforming the hypothesis space

The above analogical prediction rule used two Carnapian rules: one for R_{n+1} and one for $\bar{S}_{n+1}$ conditional on R_{n+1}. They correspond to Dirichlet priors over the hypotheses $H_{\rho\sigma}$:

$$P(R_{n+1}|H_{\rho\sigma} \cap E_n) = \rho,$$

$$P(R_{n+1}|H_{\rho\sigma} \cap R_{n+1} \cap E_n) = \sigma.$$

We can illuminate the analogical prediction rule by considering priors over these hypotheses.

Transforming the hypothesis space

We can transform the requisite prior over $H_{\rho\sigma}$ back to the prior over H_θ by using

$$\theta_A = \rho(1-\sigma), \qquad \theta_B = 1-\rho, \qquad \theta_C = \rho\sigma$$

and hence a Jacobian $1/\rho$. We obtain the following correspondence:

$$\begin{aligned}\mathrm{Dir}(\lambda\gamma_R, \lambda\gamma_{\bar{R}}) \times \mathrm{Dir}(\lambda_R\gamma_{S|R}, \lambda_R\gamma_{\bar{S}|R}) = \\ \mathrm{Dir}(\lambda_R\gamma_{\bar{S}|R}, \lambda\gamma_{\bar{R}}, \lambda_R\gamma_{S|R}) \times (\theta_A+\theta_C)^{\lambda\gamma_R-\lambda_R}.\end{aligned}$$

The term responsible for the analogical effects is $(\theta_A+\theta_C)^{\lambda\gamma_R-\lambda_R}$.

Analogy priors

The prior over $\mathcal{H}$ illuminates the analogical predictions. Priors with $\lambda_R > \lambda\gamma_R$ are warped so as to correlate high proportions of apples and of cherries.

- In the above setup, the additional factors $(\theta_A + \theta_C)$ result in a ridge along the line $\theta_A \approx \theta_C$.
- Observing a cherry will redistribute the probability over the simplex towards a higher proportion of cherries.
- As a consequence, the probability will also move towards higher proportions of apples.

Analogy priors

Owing to De Finetti's result, all exchangeable analogical predictions must somehow be encoded in a prior over $\mathcal{H}$.

- The analogical predictions of Skyrms, Paris, and others can also be illuminated by looking at the prior.
- We can define systematic relations between the above class of priors and a relevance metric among predicates.
- Clearly there are interesting analogy priors outside this class but Dirichlet priors have attractive properties.

3 Observations in attribute space

We now refine the space of predicates towards an underlying space of attributes.

The move to a richer observation space is inspired by new approaches to the notion of predicate by, e.g., Gärdenfors.

Attribute space

The space of hypotheses now doubles up as the space of attributes and hence of possible observations.

- Instead of drawing fruits from a fridge, we sample fruit juice from a blender.
- We observe proportions of fruit in each sample, $\Gamma_i = \langle \gamma_A, \gamma_B, \gamma_C \rangle_i$.
- Deviations from the true proportios arise by fallible taste or by improper mixing.
- By means of these observations we aim to determine the actual proportions H_θ in the blender.

Probabilities over attribute space

We can now employ the same machinery that we have used in the Bayesian statistical underpinning of Carnapian rules. We update with these likelihoods:

$$P(\Gamma_{n+1}|H_\theta \cap G_n) = \mathrm{Dir}(\gamma_A, \gamma_B, \gamma_C),$$

where G_n denotes the earlier observations. Starting with a uniform prior and multiplying by these Dirichlet densities, we obtain

$$P(H_\theta|G_n) = \mathrm{Dir}(n\bar{\gamma}_A, n\bar{\gamma}_B, n\bar{\gamma}_C),$$

where $\bar{\gamma}_X$ is the average over the observed γ's.

Probabilities over attribute space

The above framework naturally extends the foregoing statistical setup.

- The points in attribute space double up as hypotheses H_θ that concern the data generating system. We will briefly consider the move to a richer notion of the data-generating system towards the end.
- The posterior over attribute space converges onto the true proportions of fruits with probability 1.
- If the observation in attribute space is extremal towards, e.g., apple, $\Gamma = \langle 1, 0, 0 \rangle$, we obtain the likelihood of the apple from the statistical setup: $P(\langle 1, 0, 0 \rangle | H_\theta \cap G_n) = \theta_A$.

Predictions from attribute space

As before, we can convert the posterior probability over H_θ into a prediction of a predicate by taking an expectation value:

$$P(A_{n+1}|G_n) = \int_{\theta\in\Theta} P(H_\theta|G_n)P(\langle 1,0,0\rangle|H_\theta \cap G_n)\, d\theta.$$

The idea is that we may be forced to tick a box on the taste of the juice. Assuming a uniform prior this results in simple expressions for the predictions:

$$P(A_{n+1}|G_n) = \frac{n\bar{\gamma}_A + 1}{n+3}.$$

Predictions of categorical predicates

We can define categorical predicates like cherry-flavoured, which attach to samples, as regions in attribute space. Predictions on predicates can then be extracted from the posterior over the space:

$$P(C_{n+1}|G_n) = \int_{\theta \in \Theta_C} P(H_\theta|G_n)\, d\theta.$$

Such predictions will naturally show analogical effects. The likelihood functions are such that the occurrence of a sample within a particular predicate will benefit nearby predicates most.

Analogical predictions

We can introduce analogical effects into either version of predictions based on attribute space by tweaking the likelihood functions. The parameterisation in terms of component predicates is instrumental for this,

$$P(\Gamma_{n+1}|H_{\rho\sigma} \cap G_n) = \mathrm{Dir}(\lambda\gamma_R, \lambda\gamma_{\bar{R}}) \times \mathrm{Dir}(\lambda_R\gamma_{S|R}, \lambda_R\gamma_{\bar{S}|R}).$$

To encode a relevance between apples and cherries, we must now choose $\lambda\gamma_R > \lambda_R$: different proportions of apple and cherry are less distinguishable than different proportions of banana.

Questions, questions, questions

The idea of inductive logic over attribute spaces invites several further questions.

- The eventual predictions are exchangeable and this leaves us to wonder about the corresponding analogy prior.
- The likelihood functions express relevances among attributes. It seems attractive to define a metric over attribute space by means of these likelihoods.
- The relevances may be asymmetric: cherries may be indicative of more apples but not conversely. How can this be arranged otherwise?

4 A richer observation source

The above move towards attribute space is only a half-way house: the observations are refined but their source is not.

We might imagine that the source of the observations is itself characterised by a distribution over attribute space.

Statistical inference on attribute space

The inductive logic for attribute space now appears as a standard statistical analysis.

- We assume a model, i.e., a set of distributions over attribute space, and a prior probability over it.
- Relative to the observations we can determine a posterior probability assignment over the distributions in the model.
- Predictions on attributes and predicates can be determined by computing marginal likelihoods for the fruit juice proportions.

Analogical predictions

We can introduce analogical effects into these systems by constraining the statistical model in various ways.

- If the distributions over attribute space are unimodal, in the sense that the derivative only changes sign once, then an observation Γ will favour nearby fruit juice proportions.
- If we constrain a set of multimodal distributions in a specific way, we can also orchestrate analogical effects somewhat akin to those of hyper-Carnapian rules.

5 Analogy by similarity as proximity

We have seen two approaches to predictive systems that introduce analogical effects among the predicates. They share the idea that similarity, as a basis for the analogy between predicates, relates to proximity.

- In predicate space, the proximity is expressed by the Hamming distance in terms of the underlying predicates: apples and cherries are closer because they are both round.

- In attribute space, the proximity is the metric of the space. We can develop this by defining the metric in terms of the likelihood functions over the space, i.e., by a graded distinguishability of the hypotheses.

Inductive logic: what is next?

This seems to be a suitable moment for reflecting on the future of inductive logic.

- What makes the topic timely and important? Perhaps the development of big data research and the increased prominence of data-driven, or at least statistical methods.

- What are its areas of growth? Perhaps its relations to the cognitive sciences, to machine learning, and to evolutionary, interactive and social settings, e.g. prediction games.

Thank you

The slides for this talk will be available at http://www.philos.rug.nl/ romeyn.

For comments and questions, email j.w.romeijn@rug.nl.