Analogical Predictions

Jan Willem Romeyn, Groningen

1. Introduction

This paper deals with exchangeable analogical predictions, and proposes a Bayesian model for such predictions. The paper first discerns two kinds of analogical predictions, based on similarity of individuals and of types respectively. It then introduces a Bayesian framework that employs hypotheses for making predictions. This framework is used to describe predictions based on the similarity of individuals, and further relates exchangeable predictions with a specific partition of hypotheses on types. Exchangeable predictions based on type similarity are determined by prior probabilities over the partition, but the partition obstructs the control over the similarity relations. Finally the paper develops a model for exchangeable predictions based on type similarity, which employs hypotheses on similarity between individuals, thereby offering a better control over the similarity relations.

2. Similarity of individuals and types

This section introduces the two kinds of similarity that play a role in analogical predictions. It further describes the relation between the predicates associated with these similarities.

Following (Niiniluoto 1981), analogical predictions based on similarity between individuals have this form:

$M_1a_1 \cap M_2a_1$

$M_1a_2 \cap M_2a_2$

...

$M_1a_t \cap M_2a_t$

M_1a_{t+1}

____

probably M_2a_{t+1}.

That is, the similarity between individuals a_1 to a_t and the further individual a_{t+1} is derived from the fact that all of them satisfy the predicate M_1. This similarity is subsequently used to derive from M_2a_1 to M_2a_t that probably also M_2a_{t+1}. The similarity involved in this inference is between individuals a_1 to a_t on the one hand, and the further individual a_{t+1} on the other.

Imagine that we have a limited number L of M-predicates, indexed k, which can be either true or false of an individual. Denote the above predications of a_i as $M^m{}_k\, a_i$, where m = 0 or m = 1 for M_ka_i being true or false respectively. We can define the type predicate of an individual a_i as cells Q in the partition generated by the M-predicates. These cells are determined with an L-tuple q of the binaries m, which encode the satisfaction of M-predicates by individual a_i:

$$q = \langle m_1, m_2, \ldots , m_L\rangle .$$

So every individual a_i can be assigned a unique type predicate Q^q, which refers to one particular cell in the partition generated by the M-predicates. I sometimes refer to the type predicates as Q-predicates, and where convenient I denote the L-tuples q with natural numbers.

Imagine that we only have access to the types, or Q-predicates, and not to the M-predicates underlying these types. In that case we can still make analogical predictions. For example, it may be that we consider type Q^1 more similar to type Q^2 than type Q^0, where these type numbers refer to specific L-tuples q, and further that we observe the t-th individual to be of type Q^1. Then, apart from deeming Q^1 more probable for the next individual, we may also take the effect of observing Q^1 to be more favourable to Q^2 than to Q^0. The observation of Q^1 is thus assumed to have some predictive relevance for Q^2 as well. Below I give a formal definition of this kind of analogical prediction. For now, note that the example is an analogical prediction based on similarity between types of individuals.

3. Bayesian framework

The following contains a brief introduction into a Bayesian framework that employs hypotheses for making predictions. It deals with observations, belief states, hypotheses, updating by conditioning, and predictions.

In expressing the observations of Q- or M-predicates, it will be convenient to omit reference to the individuals a_i. Instead we can refer to the observations directly by adding a time index to the predicates. Thus, the expressions $Q^q{}_t$ and $M^m{}_{kt}$ refer to the observations of individual a_t having predicates Q^q and $M^m{}_k$ respectively. The expressions $E^Q{}_t$ and $E^{Mk}{}_T$ refer to sequences of such observations, having length t, or a vector of such lengths T. We can write

$$Q^q{}_t \cap E^Q{}_{t\text{-}1} = E^Q{}_t\,,$$

$$M^m{}_{kt} \cap E^M{}_{Tb} = E^M{}_{Ta}\,,$$

in which T_b and T_a are vectors of lengths having t-1 and t in the k-th element respectively. The remainder of this section deals with Q-predicates. The formal treatment of M-predicates is analogous.

Belief is represented with a probability function p, which takes observations and hypotheses as arguments. Observations $Q^q{}_t$ are defined above. Hypotheses H are general observational statements that prescribe specific probabilities for the observations $Q^q{}_t$. The prescribed probabilities $p(\, Q^q{}_t \mid H \cap E^Q{}_{t\text{-}1})$ are called the likelihoods of observations $Q^q{}_t$ on hypothesis H. A partition of hypotheses $P = \{\, H_\alpha \mid \alpha \in [0,1]\, \}$ is a collection of such general statements that are together assumed to exhaust logical space.

When we observe $Q^q{}_t$, our new belief can be represented with the same probability function conditioned on the new observation:

$$p(\,\cdot \mid E^Q{}_{t\text{-}1}) \quad\rightarrow\quad p(\,\cdot \mid Q^q{}_t \cap E^Q{}_{t\text{-}1})\,.$$

This transition between belief states is called updating by conditioning. We can update the probability assigned to hypotheses by means of the likelihoods on these hypotheses:

$$p(\, H_\alpha \mid Q^q{}_t \cap E^Q{}_{t\text{-}1}) = p(\, H_\alpha \mid E^Q{}_{t\text{-}1}) \times p(\, Q^q{}_t \mid H_\alpha \cap E^Q{}_{t\text{-}1}) \,/\, p(\, Q^q{}_t \mid E^Q{}_{t\text{-}1})\,.$$

Moreover, using the partition we can write the denominator $p(\, Q^q_t \mid E^Q_{t\text{-}1})$ in terms of likelihoods and probabilities over hypotheses:

$$p(\, Q^q_t \mid E^Q_{t\text{-}1}) = \int p(\, H_\alpha \mid E^Q_{t\text{-}1}) \times p(\, Q^q_t \mid H_\alpha \cap E^Q_{t\text{-}1})\, d\alpha \, .$$

The predictions $p(\, Q^q_{t+1} \mid E^Q_t\,)$ can be expressed in the same way. In sum, they can be derived from a partition P, a prior probability $p(H_\alpha)$, the likelihoods $p(\, Q^q_i \mid H_\alpha \cap E^Q_{i\text{-}1})$ for i ranging from 1 to t+1, and the observations E^Q_t.

The above may seem an unnecessarily complicated framework for making predictions: we can also define predictions directly, as a function of previous observations E^Q_t and some further parameters. However, as I will show, hypotheses are very useful in laying down the assumptions underlying analogical predictions.

4. Similarity between individuals

Consider again the analogical predictions based on similarity between individuals, as introduced in section 1. We can model these predictions using the framework of section 2. This section defines the model by means of a specific partition of hypotheses.

To model similarity between individuals we can use a partition of hypotheses H_μ , which is summarised with the following diagram:

$M_1 \downarrow$ / $M_2 \rightarrow$	1	0
1	$\mu_1\mu_2^1$	$\mu_1(1\text{-}\,\mu_2^1)$
0	$(1\text{-}\,\mu_1)\,\mu_2^0$	$(1\text{-}\,\mu_1)(1\text{-}\,\mu_2^0)$

Every hypothesis H_μ is characterised by a vector μ containing parameters μ_1, μ_2^1, and μ_2^0. The diagram shows the likelihoods of the M-predicates conditional on these hypotheses. The parameter μ_1 is the likelihood of M^1_{1t} for any t:

$$p(\, M^1_{1t} \mid H_\mu \cap E^M_{\langle t\text{-}1,t'\rangle}) = \mu_1 \, .$$

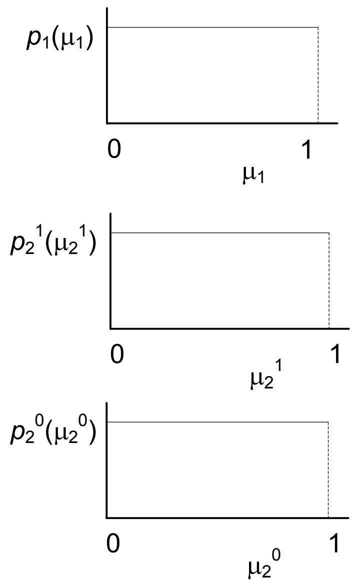
The parameters μ_2^1 and μ_2^0 are the likelihoods of M^1_{2t} on H_μ , conditioned on the further occurrence of M^1_{1t} and M^0_{1t} respectively:

$$p(\, M^1_{2t} \mid H_\mu \cap E^M_{\langle t\text{-}1,t'\rangle} \cap M^1_{1t}) = \mu_2^1 \, ,$$

$$p(\, M^1_{2t} \mid H_\mu \cap E^M_{\langle t\text{-}1,t'\rangle} \cap M^0_{1t}) = \mu_2^0 \, .$$

The probability distribution over the partition of hypotheses H_μ is a function of the three independent parameters in μ.

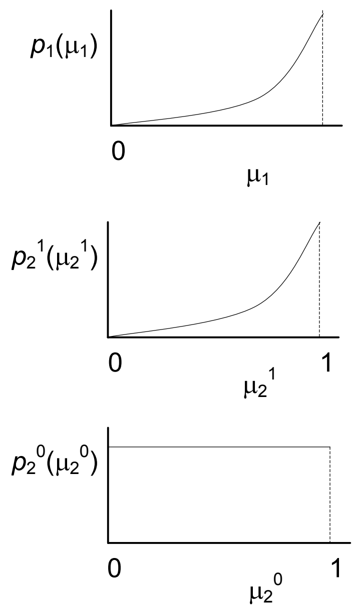
With this partition of hypotheses in place, I can make explicit how the predictions resulting from it employ similarity between individuals. Assume that the prior probability distribution over hypotheses H_μ is uniform, so that the marginal distributions over the parameters in μ are uniform too:



Furthermore, assume that the observations until t are those given in the example of section 1:

$$E^M_T = M^1_{11} \cap M^1_{21} \cap M^1_{12} \cap M^1_{22} \cdots M^1_{1t} \cap M^1_{2t} \, .$$

We can then compute a probability over the hypotheses conditional on these observations. The marginal distributions of this updated distribution will look approximately as follows:



Since there has been no relevant observation on the value of μ_2^0, the distribution over μ_2^0 has not changed. But the distributions over μ_1 and μ_2^1 are both tilted towards higher values. This is because updating with the observations of M^1_{1i} and M^1_{2i} favours hypotheses that have higher likelihoods μ_1 and μ_2^1 for these observations.

Now we can compare the prediction of M^1_{2t+1} based on the observations E^M_T and $E^M_T \cap M^1_{1t+1}$ respectively. This difference measures the effect of taking individual a_{t+1} to be similar to the observed individuals a_1 to a_t, for which M_1 was true. Using the above likelihoods, the predictions are

$$p(M^1_{2t+1} \mid E^M_T) = \int p(H_\mu \mid E^M_T) \times [\mu_1\mu_2^1 + (1-\mu_1)\,\mu_2^0]\, d\mu \;,$$

$$p(M^1_{2t+1} \mid E^M_T \cap M^1_{1t+1}) = \int p(H_\mu \mid E^M_T) \times \mu_2^1\, d\mu \;.$$

It can easily be verified that the prediction based on E^M_T and the further observation that M^1_{1t+1} is higher than the one based on E^M_T alone. The Bayesian framework thus models analogical predictions based on similarity between individuals.

5. Exchangeability and type similarity priors

The above concerns predictions based on direct observations of *M*-predicates. Similarity of individuals is exhibited in the *M*-predicates that the individuals have in common. However, in many cases we do not have access to *M*-predicates, but only to *Q*-predicates. Similarity between individuals must then be replaced by similarity between types. The remainder of this paper is devoted to analogical predictions based on type similarity.

The following only discusses exchangeable analogical predictions. Let the operation $\Phi_i[\,\cdot\,]$ permute the first and the *i*-th observation of a sequence E^Q_t. The exchangeability of a prediction can then be defined as

$$p(Q^q_{t+1} \mid \Phi_i[E^Q_t]) = p(Q^q_{t+1} \mid E^Q_t) \;.$$

This means that the prediction does not depend on the order of the observations Q^q in E^Q_t. In a Bayesian framework, exchangeable predictions can be represented with a specific partition of hypotheses. Assuming that the types *q* are numbered 0 to *K* and indexed *j*, these hypotheses have the following likelihoods:

$$p(Q^j_t \mid H_\alpha \cap E^Q_t) = \alpha_j \;.$$

A hypothesis H_α is thus defined by a vector α containing parameters α_0 to $\alpha_{K\text{-}1}$. The parameter α_K must be such that $\Sigma_j\alpha_j = 1$.

Let me now give a formal definition of type similarity. Following Carnap's original suggestions in (1980), the similarity effect may be rephrased as

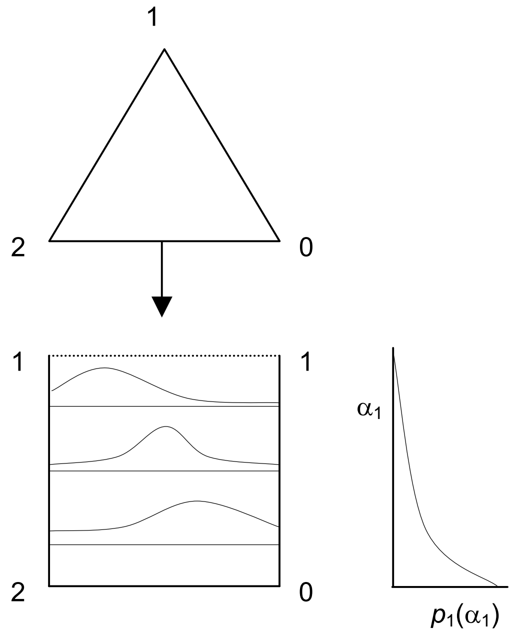
$$s_{12} > s_{10} \Rightarrow p(Q^1_{t+1} \mid Q^2_t \cap E^Q_{t\text{-}1}) > p(Q^1_{t+1} \mid Q^0_t \cap E^Q_{t\text{-}1}) \;,$$

where s_{ij} denotes the similarity between types *i* and *j*. For the purpose of this paper, I take similarity to be reflexive, so that from $s_{12} > s_{10}$ we also have

$$p(Q^2_{t+1} \mid Q^1_t \cap E^Q_{t\text{-}1}) \,/\, p(Q^2_{t+1} \mid E^Q_{t\text{-}1}) \;>$$
$$p(Q^0_{t+1} \mid Q^1_t \cap E^Q_{t\text{-}1}) \,/\, p(Q^0_{t+1} \mid E^Q_{t\text{-}1}).$$

This is a formal expression of the type similarity in the example of section 1. I will not discuss the improved and refined definitions of type similarity suggested in (Kuipers 1984a) or (Festa 1997).

Exchangeable analogical predictions can now be characterised using the partition of H_α, and an appropriate prior over this partition. Intuitively, it must look somewhat like the following:



The simplex on the left is first transformed into a square by stretching the top corner. The prior over the simplex for H_α can then be decomposed in a marginal distribution over α_1 and a continuum of distributions over the transformed simplex, each of them associated with a specific value of α_1. The idea behind the continuum is that higher values of α_1 are associated with distributions that favour Q^2 over Q^0. If we observe Q^1, the marginal distribution over α_1 will tilt towards the higher values, and this will cause the distributions favouring Q^2 to gain more weight. The marginal distribution for α_0 will then shift slightly towards lower values.

However, this representation makes it difficult to translate the update with Q^1 into an operation on the marginal distribution for α_0 or α_2. A related difficulty is in encoding type similarity relations into the priors. Skyrms (1993), Maher (2001) and others have presented interesting ways to do this more or less indirectly. The last section develops a new method, based on the model of section 3.

6. Type similarity model

The idea behind the model is that similarity relations between types can be expressed in terms of shared underlying *M*-predicates. This section employs the hypotheses of section 3 to make this idea precise.

Following the example, types Q^1 and Q^2 are highly similar, and both are dissimilar to Q^0. We can now employ the hypotheses H_μ to facilitate an expression of the similarity relations. The diagram represents the new hypotheses $H_{\alpha\mu}$:

M_1 ↓ \ M_2 →	1	0
1	1	∅
0	2	0

This diagram gives the following relations between *M*- and *Q*-predicates:

$M^0_{1t} \cap M^0_{2t} = Q^0_t$,

$M^1_{1t} \cap M^1_{2t} = Q^1_t$,

$M^0_{1t} \cap M^1_{2t} = Q^2_t$,

$M^1_{1t} \cap M^0_{2t} = \varnothing$.

From this last relation we know that $\mu_2^1 = 0$. So the hypotheses $H_{\alpha\mu}$ are defined by only two parameters. These are related to the original likelihoods of the *Q*-predicates on the hypotheses H_α:

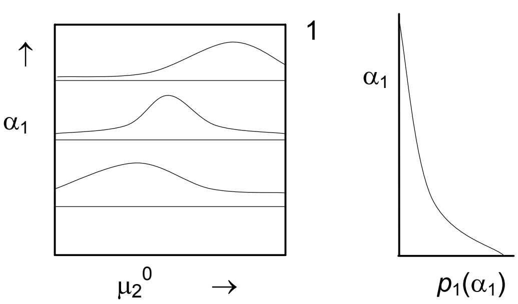
$p(Q^1_t \,|\, H_{\alpha\mu} \cap E^Q_t) = \mu_1 = \alpha_1$,

$p(Q^2_t \,|\, H_{\alpha\mu} \cap E^Q_t) = (1 - \mu_1)\, \mu_2^0 = (1 - \alpha_1)\, \mu_2^0$,

$p(Q^0_t \,|\, H_{\alpha\mu} \cap E^Q_t) = (1 - \mu_1)\, (1 - \mu_2^0) = (1 - \alpha_1)\, (1 - \mu_2^0)$.

In the following, I take the hypotheses to be parameterised with α_1 and μ_2^0.

The similarity relations between types can now be expressed with a prior over these new parameters. We can use a marginal distribution $p_1(\alpha_1|E^Q_{t-1})$, and a continuum of distributions over μ_2^0 associated with the values of α_1:



In this representation the prior is related to the similarity relations much more directly. Because the framework involves no transformation, the marginal distributions for α_j can be computed quite easily. The shift in the probability of Q^2_{t+1} due to the occurrence of Q^1_t can therefore be readily expressed:

$p(Q^2_{t+1} \,|\, E^Q_{t-1}) = \int p_1(\alpha_1 \,|\, E^Q_{t-1}) \times \alpha_1 A_\mu(\alpha_1)\, d\alpha_1$,

$p(Q^2_{t+1} \,|\, Q^1_t \cap E^Q_{t-1}) = \int \alpha_1\, p_1(\alpha_1 \,|\, E^Q_{t-1}) \times \alpha_1 A_\mu(\alpha_1)\, d\alpha_1$,

where $A_\mu(\alpha_1)$ is the average of μ_2^0 as a function of α_1. With these expressions for the shifts we have full control over the similarity relations encoded by the prior.

The above presents one example of encoding similarity between types in a prior over hypotheses that employ underlying predicates. Many more underlying predicate structures may be investigated, all of them facilitating different similarity relations. Positing such structures turns out to be a convenient tool for controlling the similarity assumptions used in analogical predictions.

Literature

Carnap, R. 1980 “A Basic System of Inductive Logic, Part II”, in Jeffrey, R.C. (ed.), *Studies in Inductive Logic and Probability* Vol.2, Berkeley: University of California Press, 7-150.

Festa, R. 1997 “Analogy and Exchangeability in Predictive Inferences”, *Erkenntnis* 45, 89-112.

Kuipers, T.A.F.: 1984a ‘Two Types of Inductive Analogy by Similarity’, *Erkenntnis* 21, 63-87.

Maher, P. 2000 ‘Probabilities for Two Properties’, *Erkenntnis* 52, 63-91.

Niiniluoto, I. 1981 ‘Analogy and Inductive Logic’, *Erkenntnis* 21, 1-34.

Skyrms, B. 1993 ‘Analogy by Similarity in Hyper-Carnapian Inductive Logic’, in J. Earman, A.I. Janis, G. Massey, and N. Rescher (eds.), *Philosophical Problems of the Internal and External Worlds*, Pittsburgh: University of Pittsburgh Press, pp. 273-282.