EMPIRICAL PREDICTION ALGORITHM

N.G. Zagoruiko
Institute of Mathematics
Academy of Sciences, USSR, Siberian Branch
Novosibirsk, 650090, USSR

Abstract

The following requirements to empirical prediction algorithms are investigated: universality, non-triviality, and consistence. It is pointed out these requirements are too strong. Some variants of algorithms in which the requirement of "consistance" is replaced by that of "the greatest simplisity" are tested.

Analysis of process of problem solving tasks by the systems, which investigators call intellectual ones, shows that stage of prediction of new facts by using regularities discovered on set of well-known facts, plays an important part in this process.

Thus, in pattern recognition tasks a belonging of control realization to one or another image is predicted on the grounds of natural regularities observed on training sequence between objects' properties and pattern name.

In chess and draughts programs, etc. a prediction block is represented by procedures of taking a decision about the most preferable move. In so doing the regularities introduced by programmer or discovered by computer are usedj the winning move is that one when estimation function of such-and-such kind reaches the maximum value; in such-and-such situation from the list preliminarily "learnt by heart" such-and-such move is successful, etc.

In programs intended for definition of structural formulas of organic molecules by their chemical formula and by mass-spectrum, the mass-spectra of molecules engendered by hypothesis generator are calculated with the aid of prediction block. Results of this prediction are used for selection the structures which should have mass-spectrum not so much different from experimental one.

Ability of making successful predictions is one of the most important features of any intellectual system. One can attain a certain object only in case when there is a possibility to foresee consequences of one or another actions. On this subject von Foerster writes that to survive is to foresee correctly the events in surroundings. Inductive inference is the logical basis of foresight, i.e. method to search, given condition E, for hypothesis h, which is confirmed by surroundings S and is convenient for a certain aim [1].

In the last analysis, purpose of any empirical science is to bring to light the natural regularities between characteristics of observable phenomena, and to formulate these regularities in the form that could be reliable and convenient means to foresee the new phenomena.

Pact of foresight (prediction) can be established objectively and definetly. Description of the event to be predicted is entered in protocol; correctness of the prediction is tested by course of further events.

Especial importance of the stage "prediction" for intellectual systems and the possibility of constructive definition of the stage justifies concentration of efforts to its study. This paper contains review of works on methodology and of algorithms of prediction of facts and events in empirical world.

Possibility to predict events is based on acceptance of determinism conception. Denial of causal relation between phenomena automatically excludes such possibility. One can illustrate negative position relative to any prediction by Wittgenstein's opinion [2] that we cannot predict events of the future on the basis of the ptesent. Belief in causality is prejudice. That the sun will rise tomorrow is a hypothesis; in other words, we do not know firm, whether it will rise.

In the following we shall proceed from the belief in existence of natural relations between phenomena. Predictions will be made with application of regularities that are found in empirical data. Discussion of necessary requirements for methods (sometimes, algorithms) of prediction will be the main aim of our work.

Let us consider a number of concepts [5]. "Empirical hypothesis" we shall mean as a set of formalized notions about characteristics of objects or phenomena under study. One can speak about properties of the real world only by fixing instruments (P) which measure these properties. Let a set of symbols $O=\{P_1, R_2..., P_n\}$ used for designation of instruments tnore exactly, empirical relations measured with the aid of these instruments) be signature of hypothesis.

Let empirical interpretation of relations measured with the aid of fixed collection of instruments, designated by symbols of 6, be intenslonal basis (Int) of hypothesis.

$$Int = \{\mathcal{P}_1, \mathcal{P}_2 \dots \mathcal{P}_n\}$$

Let $\mathcal{P}_1$, for instance, describes characteristics of the instrument fixing partial ordering relation over the set of objects of $\mathcal{B}_0\{a, b, c \dots\}$

$$\mathcal{P}_1 : P_1 = 1, \ iff \ a \geqslant b:$$
$$P_1(a,b) \wedge P_1(b,c) \rightarrow P_1(a,c);$$
$$P_1(a,b) \wedge P_1(b,a) \rightarrow a = b;$$

For such instrument as ampermeter, the intensional basis will involve, for example, such expressions:

$$\mathcal{P}_2 : P_2 = 1, \ iff \ Ia + Ib \geqslant Ic;$$
$$P_2(Ia, Ib, Ic) \wedge P_2(Ia, Ib, Id) \rightarrow Id \leqslant Ic$$

Our knowledge about dependence of readings of several instruments can be, for example, put down in Int. as follows:

$$\mathcal{P}_3 : \quad P_3 = 1, \text{ iff } I_a \cdot R_a = U_a ;$$

Notions about experimental results which can be obtained with the help of these instruments or which cannot, are formalized with the aid of <u>test algorithm (T)</u>.

Empirical hypothesis can be formally represented in the form of ternary

$$H = \langle \sigma , \text{Int}, T \rangle \quad .$$

If on some set of objects (B) m-ary empirical relation is true then this is designated by symbol P_i ; and if that is false then by symbol $\bar{P}_i$; and if that is meaningless then by symbol $\tilde{P}_i$. Let sequence of symbols from σ with their truth values be <u>protocol (Pr)</u>. Algorithm T shows relative to any protocol Pr whether the given protocol conforms to hypothesis H (in which case T(Pr) = 1) or refutes it (T(Pr) = 0).

Let, for instance, the weight of objects a and b is equal to 5 and 2 kg., accordingly. Then, the protocol representing the result of comparison of a and b in weight by instrument $\mathcal{P}_1$ will look as follows:

$$Pr' = \{P_1(a,a), P_1(b,b), P_1(a,b), \bar{P}_1(b,a)\}.$$

Such a protocol does not contain contradictions from the view-point of characteristics of the instruments $\mathcal{P}_1$ and so T(Pr')=1. But if for some pair of objects c and d we shall write

$$Pr'' = \{P_1(c,c), P_1(d,d), \bar{P}_1(c,d), \bar{P}_1(d,c)\}$$

such a protocol would contradict characteristics of instrument $\mathcal{P}_1$: d cannot be heavier than c and simultaneously c cannot be heavier than d . As the result the protocol Pr' would be rejected as incorrect one, i.e.

$$T(Pr'') = 0$$

Pair consisting of empirical hypothesis H and of protocol Pr, conforming to hypothesis H, is <u>admissible pair</u> H,Pr .

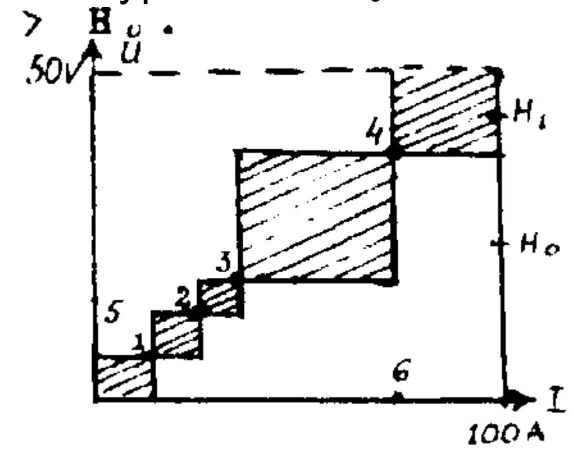
We proceed from conviction that without essential restriction of generality a distinct act of empirical prediction may be considered as follows. Originally there is a protocol (a record) Pr_0 of experiment over finite set B_0 of empirical objects. This protocol is considered as a mere registration of results of interaction of these objects with instruments used. Generally speaking, there also is an empirical hypothesis H_0 affirming that certain protocols can be never obtained if H_0 is true, and if the experiments are made with the given instruments over any finite sets of objects (and not only over the set B_0). The hypothesis H_0 may be regarded as a description of supposed characteristics of measuring instruments, and the protocol Pr as a record of results of measurements of elements of B_0 carried out with these instruments* The hypothesis H_0 is supposed to be such that protocol Pr_0 , corresponding to the experiment made over the set B_0 is admissible under this hypothesis (i.e. the hypothesis H_0 conforms to the protocol Pr_0). Otherwise we must state the hypothesis is refuted by this experiment and it must be revised as a wrong initial information. It is precisely the case when our initial assumptions concerning characteristics of measuring instruments are wrong.

A distinct act of prediction, say $\langle H_0, Pr_0, B_0 \rangle \rightarrow H_1$, is that starting from the initial hypothesis H_0 and using information involved in the protocol Pr_0 concerning elements of B_0 a new hypothesis H_1 is pointed out, such that:

(i) H_1 is in a sense more (or at least not less) informative than H_0 ;

(ii) Pr_0 is admissible for the H_1 .

Let from the first we know only that current in circuit may have any value from 0 to 100 amperage, and tension of this circuit is from 0 to 50 volt. Then hypothesis H_0 would regard any combination of mentioned values of current and tension as possible one. As a result of experiment over some part of electric circuit there have been obtained such combinations: 1. (U=3v, I=6a)t 2. (U=10v, I=20a); 3. (U=15v, I=30a); 4. (U=40v, I=80a). Using these results one can construct the hypothesis H_1 interdicting, in contrast to H_0 .such pair, for example, as follows: 5. (U=5v, I=1a)| 6. (U=0v, I=90a), etc. One can express in figure this situation. In such a case we shall consider the hypothesis H_1 to be stronger than the hypothesis E_0 • and write this as $H_1 > H_0$.



The act of prediction is considered as successful (or true) till a new set of empirical objects, such that protocol of experiment of kind mentioned over this set is not admissible for H_1 and is admissible for H_0 , is found. It is obvious, trivial act of prediction, i.e. any act of type $\langle H_0, Pr_0, B_0 \rangle \rightarrow H_0$, is always successful in this sense, being completely uninteresting.

An individual act of prediction has so far been concerned. As for method of prediction it is natural to consider it as a function f of type:

$$f(\langle H_0, Pr_0, B_0 \rangle) = H_1 .$$

Of course, certain restrictions that arise from our intention to impart certain desirable features to prediction method should be placed upon this function f.

In [4] the following formulas of these requirements were given.

1. Universality.

$$\forall \langle H_0 Pr_0 \rangle \exists H_1 (f(\langle H_0 Pr_0 \rangle) H_1).$$

The sense of this requirement is evident: the algorithm would be applicable to any possible pair "protocol-hypothesis",

2. Non-triviality.

$$H_1 \gg H_0 .$$

Hypothesis resulting from algorithm work would be, at least, no weaker than the initial one.

3. Consistency.

Let φ be one-one computable transformation of possible pair $\langle H_0, Pr_0 \rangle$ into possible pair $\langle H_0', Pr_0' \rangle \varphi = (\langle H_0, Pr_0 \rangle)$, such that hypotheses H_0 and H_0' are refuted or confirmed on the same sets of objects (i.e. simultaneously). In such a case we can regard φ as effective one-one translation of hypothesis H_0 and protocol Pr_0 into another equivalent language. It is natural to demand invariance of results of prediction relative to such effective one-one translations from one language into another equivalent one; i.e. if $f\langle H, Pr_0 \rangle = H_1$, and $f\langle H_0', Pr_0' \rangle = H_1'$, then H_1 and H_1' would be refuted or confirmed simultaneously.

K.F. Samochvalov has proved the theorem 4 that the only function answering these three requirements is the function f• constructing the decoder, i.e. function that brings into correlation pair $\langle H_0, Pr_0$ with hypothesis H , interdicting all protocols except Pr corresponding to training sequence. It is clear, such a function f* cannot be means of inductive generalization or that of discovery of empirical laws. To obtain useful rediction method the requirements formuated above must be changed. Very likely, we have to reduce the third requirement of invariance of prediction with respect to formally equivalent ways of initial data representation. For this reason attempts to construct universal, non-trivial, and useful method of prediction presuppose acceptance of "Goodman's approach" [5] : one ought to prefer some languages to another ones if, in spite of their equivalence, they differ on such criteria as "habitualness" and "frequency of usage" of the terms.

The work [4] , very likely, may be a justification of widespread in science principle of "simplicity" [6,7] . According to this principle, from two theories equally well-conformed to known fact one ought to use more "simple" one to predict new events.

Along with research of requirements to empirical prediction algorithms, some variants of algorithms in which the requirement of "consistance" is replaced by that of "the greatest simplisity" axe tested. One of much algorithms is as folioing [8] :

1. Let Pr_0 be the protocol of experiment over the finite set B_0 of empirical objects.

Let H_0 be the initial hypothesis.

Suppose we are interested in prediction concerning the ℓ new objects on the grounds of more strong hypothesis H . Such prediction is equivalent to choice of one protocol Pr,, having cardinal $n+\ell$ from all protocols having the same cardinal, which conform to hypothesis H_1

We usually know something about new empirical objects. We want to predict the values of some empirical relations P_j Let subprotocol describing the set of these relations be E :

$$E = \{Pr_j\}.$$

2. To predict subprotocol E we construct (generate) all protocols $\{Pr_1\}$ having cardinal $n+\ell$ such that:

a) each of them conforms to hypothesis H_0

b) for each of them the set E is its own subset.

3. We shorten obtained list of protocols as follows:

a) For every protocol $Pr_1^{(i)}$ from the list $\{Pr_1\}$ we find the subprotocol pr that contains all non-isomorphic subprotocols having cardinal K. (Number K is equal to the number of calling the point 3, so that, for example, in the first time K=1. in the second K=2, and so on.)

b; We choose the subprotocols $pr_1^{(i)}$ having minimum cardinal, i.e. $K=n_{min}$.

c) We eliminate from the list $\{Pr_1\}$ (and from further discussion) all those protocols Pr the subprotocols pr of which have cardinal more than n_{min}

4. If in carrying out the step 2 we eliminated from the list if only one protocol, and if remaining list consists of more than one protocol, we pass on to recarrying out of step J.

5. The remaining set of protocols together with all isomorphic to them corresponds to the hypothesis H sought.

Clearly, the step 3 provides the requirement of non-triviality for the algorithm. At the same time it is the formalization of the hypothesis of "simplicity" in accordance with which from two rival hypotheses the one that is based on regularity, observable on the lesser number of objects, will predict more "successfully".

The examination of algorithms of such types was carried out on the tasks of discovering regularities in empirical tables of different kinds. The examples of such problems follow.

"The Ohm's law", **Signature** σ contains predicate symbols $P_v^{(1)}, P_I^{(1)}, P_R^{(2)}, P_v^{(2)}, P_I^{(3)}, P_R^{(3)}$. Their interpretation is as follows.

$P^{(2)}_{V}$ symbolizes 2-ary relation "greater or equal" tested on any two readings of voltmeter; $P^{(3)}_{V}$ fixes 3-ary relation (of type $a+b \geqslant c$) tested on any three readings of voltmeter. Symbols P_I and P_R mean analogous relations on the readings of ampermeter and ohmmeter. Test algorithm of initial hypothesis H_0 consideres as admissible any finite values of current, resistance, and tension.

Table 1.

X_I	X_R	X_V	
66	8	528	
83	7	581	
90	19	1260	
78	23	1794	
59	26	1534	
72	34	2448	
87	35	3045	
61	41	2501	Pr_2
72	52	3744	
84	53	4452	
97	48	4656	Pr_1
56	56	3136	
74	64	4736	
64	71	4544	
73	110	7488 - 9472* (8030)	
72 -112* (90)	90	8100	

Protocol Pr_0 results from table 1, containing the readings of ampermeter (X_I), ohmmeter (X_R), and voltmeter (X_V). Table with predicted values of omitted elements is protocol Pr_1. (There are readings of instruments in brackets)

"Mendel's law". Signature contains symbols

$P^{(2)}_{\text{нп}}, P^{(3)}_{\text{нп}}, P^{(4)}_{\text{нп}}, P^{(2)}_{\text{р}}, P^{(3)}_{\text{р}}, P^{(4)}_{\text{р}}, P^{(2)}_{\text{г}}, P^{(3)}_{\text{г}}, P^{(4)}_{\text{г}}$

that are interpreted as 2-ary, 3-ary, and 4-ary relations (of type $a \geqslant b$, $a+b$ c, $a+b+c \geqslant d$) on the set of instrument readings which measure the number of accumulation ($P_{\text{нп}}$), quantity of rosy flowers ($P_{\text{р}}$), and quantity of blue flowers ($P_{\text{г}}$). Test algorithm of hypothesis H admits integer values in table 2, which is the good grounds for construction of protocol Pr_0. The predicted values are ticked off; there are real values of quantities sought in brackets.

Table 2.

$X_{\text{нп}}$	$X_{\text{р}}$	$X_{\text{г}}$
0	1024	0
1	768	256
2	640	384
3	576	448
4	540- 546* (544)	480
5	528	496
6	520	504
7	516	508

"Two-dimensional numeral table". The task consists in discovering regularities between numbers of lines (y_i), numbers of columns (x_j), and elements z_{ij}, being at the intersection of these lines and columns in the table 3.*)

Table 3.

	0	1	2	3	4	5	6	7	8	9
0	1	2	2	3	4	5	6	7	8	9
1	1	2	3	4	5	6	7	8	9	10
2	2	3	4	5	6	7	8	9	10	11
3	3	4	5	6	7	8	9	10	11	12
4	4	5	6	7	8	9	10	11	12	13
5	5	6	7	8	9	5	5	5	5	5
6	6	7	8	9	10	5	5	5	5	5
7	7	8	9	10	11	5	5	5	5	5
8	8	9	10	11	12	5	5	5	5	5
9	9	10	11	12	13	5	5	5	5	5

This hypothesis describes features of table 3 briefly enough and precisely (with not great redundance only;.

Theoretical and experimental investigations of empirical prediction algorithms are still in progress.

References.

1. Von Foerster H., Inselberg A., Weston, Memory and Inductive inference, in: Cybernetic Problems in Bionics, ed. by H.L. Oestreicner ana D.R. Moore, Gordon a Breach Sci. Publ., New York-London-Paris, 1968, pp.31.
2. Wittgenstein L., Tractatus Logico-Philosophicus- New York, 1963 proposition 5 1361.
3. Витяев Е.Е., Гаврилко Б.П., Загоруйко Н.Г., Самохвалов К.Ф. Требования к алгоритмам эмпирического предсказания, Вычислительные системы, 50, стр. 100-105, 1972.
4. Самохвалов К.Ф. О теории эмпирического предсказания, Вычислительные системы, 55, стр. 3-32, 1973.
5. Goodman N., Fact, Fiction, and Forecast, sec. ed., The Bobbs-Merrill Company, Inc., Indianapolis, New York, Kansas City, 1965.
6. Zagoruiko N.G., Hypotheses of the simplicity in the pattern recognition, Sec. Int. Conf. on Artificial Intellegence, London, 1972, pp. 318-321.
7. Загоруйко Н.Г. Методы распознавания и их применения. М., Советское радио, 1972.
8. Гаврилко Б.П., Загоруйко Н.Г. Алгоритм эмпирического предсказания, Вычислительные системы, 55,стр.134-138, 1973.