AUTOMATIC PREDICTION OF THE COURSE OF DISEASES WITH THE AID OF A DIGITAL COMPUTER

1969

SovietRxiv

Abstract

Full Text

UDC 51:62-506.2

CYBERNETICS AND CONTROL THEORY

S. Ya. ZASLAVSKII, K. A. IVANOV-MUROMSKII

AUTOMATIC PREDICTION OF THE COURSE OF DISEASES WITH THE AID OF A DIGITAL COMPUTER

(Presented by Academician V. M. Glushkov on 30 XII 1968)

Known methods for predicting the outcome of certain classes of diseases (see, for example, ([1,2])) are based on attempts to model a specific disease, using notions about its mechanisms. However, the mechanisms of many diseases are still unclear. Moreover, since the course of diseases depends substantially on the norm, one can speak only of types of models corresponding to definite types of constitution.

In the present article a general method of prediction is considered, based on a simplified scheme of a physician' s reasoning. The initial information contained in case histories is described in substantive terms and in this form cannot be used in a digital computer. The concept of a state is introduced, understood as a set of signs, complications, elements of medical technique, etc., determining to some extent the appearance of other states. Then the course of a disease can be approximately described by a series of successive states. In contrast to ([1,2]), what is essential here is the indication of intermediate states leading to one or another outcome.

The states are numbered, and all possible variants of the course of a fixed disease, represented by sequences of numbers of states, are entered into the memory of the digital computer. On the basis of a statistical analysis of the initial information, the probabilities p_{ij} of transition from state i to j are computed. Thus, the memory of the digital computer stores a finite directed random graph G corresponding to the given class of diseases. A query is represented by a Boolean vector $\mathbf{v} = (v_1, \dots, v_K)$, where K is the number of vertices of the graph G, and the components equal to one correspond to states observed in a particular patient or essential for his vital activity. The prediction is also represented by a Boolean vector $\mathbf{w} = (w_1, \dots, w_K)$, in which the components equal to one correspond to states determining the most probable course of the disease.

The procedure for constructing the prediction $\mathbf{w}$ for a fixed query $\mathbf{v}$ consists of two stages: preliminary selection of information from G for the given $\mathbf{v}$ and construction of $\mathbf{w}$ on the basis of the selected information.

To each vertex x of the graph G there is assigned a pair (s_x, π_x) of real non-negative numbers. The quantities s_x characterize the significance of the state in prediction and are determined by polling specialists. The quantities π_x are initially assumed to be equal to zero. Two operating modes are considered: an examination mode, determined by a sequence of the form query—prediction, and a training mode, determined by a sequence of the form query—prediction —evaluation ([3]). In this case the evaluation is carried out both by the digital computer itself, by changing the quantities s_x, π_x, and by the physician.

Let x and y be, respectively, the smallest and largest of the numbers of the components of the vector $\mathbf{v}$ that are equal to one. Then by $l_G^{\mathbf{v}}(x, y)$ we denote the set of all vertices of the graph G that belong to at least one path connecting the vertices corresponding to x and y, and containing all vertices j such that $v_j = 1$. We define the function $\psi_u^{\mathbf{v}}$ for each vertex

u $(u \in G)$ the relation

$$\psi_u^v = \sum_{z \in l_G^v(x,y)} \omega_{zu} s_z - \sum_{z \notin l_G^v(x,y)} \omega_{zu} s_z,$$

where

$$\omega_{zu} = \begin{cases} 1 & \text{with probability } p_{zu}, \\ 0 & \text{with probability } 1 - p_{zu}; \end{cases}$$

p_{zu} is the probability that z and u are, respectively, the beginning and the end of an arc in G.

Preliminary selection of information is carried out by means of a sequence of functions $\{e_i(x)\}_{1 \le i \le K}$. Each function associates with a certain vertex x of the graph G a set of vertices that are ends of arcs whose beginning is x, and that satisfy the inequality $\psi_u^v > \pi_u$. The domain of definition of $e_1(x)$ is the set of vertices for which $v_x = 1$. The function $e_i(x)$ $(i > 1)$ is defined on the set of vertices that are values of the function $e_{i-1}(x)$. The result of the selection is the graph T_v, containing the vertices belonging to the union of the sets of values of the functions $\{e_i(x)\}$, and the arcs are the same as in G. If T_v contains not a single path to which all x $(v_x = 1)$ belong, then we shall say that, in response to the query $\mathbf{v}$, no prognosis is issued.

Suppose that from each vertex x a transition is possible to no more than two vertices y and z. Let us associate with the pair of vertices (x, y) a random variable $b_{xy}^v(N)$, depending on the method of joining the vertices of the training sequence (t.s.), as well as on the value of the difference $\psi_y^v - \pi_y$ (4). Denote $\Delta B = b_{xy}^v(N) - b_{xz}^v(N)$. The construction of the prognosis is realized by the operator Q, which is specified by the equality $Q(x) = \Delta B$. This means that for $\Delta B \ge 0$ the vertex y $(w_y = 1)$ belongs to the prognosis, and for $\Delta B < 0$ the vertex z $(w_z = 1)$ does. The case when from the vertex x a transition to

n vertices is possible reduces to the preceding one by the method of successive dichotomy.

The notation $l_G^v(x,y) \xrightarrow{m \to z}$ ($z \xrightarrow{m} l_G^v(x,y)$) will be understood as the set of vertices of the graph G that are the end (beginning) of an arc whose beginning (end) belongs to $l_G^v(x,y)$. Let the t.s. L' be obtained from L by adjoining the vector $\mathbf{v}$: $L' = L\mathbf{v}$. Let $s_x(L)$ be the value of the quantity s_x after training by means of the t.s. L. Then for $z \in T_v$ and a given constant C

$$s_z(L') = \begin{cases} s_z(L)+1, & \text{if } z \in l_G^v(x,y),\ s_z(L) \neq C, \\ s_z(L)-1, & \text{if } z\circ \to l_G^v(x,y) \text{ or } l_G^v(x,y)\cap \to z,\ s_z(L) \neq 0, \\ s_z(L) & \text{in all other cases.} \end{cases} \tag{1}$$

If π_i changes during training, then the laws of change can be specified in the form

$$\pi_z(L') = \begin{cases} \pi_z(L), & \text{if } i \in l_G^v(x,y), \\ \pi_z(L)+1 & \text{otherwise} \end{cases} \tag{2}$$

or

$$\pi_z(L') = \begin{cases} \pi_z(L)+1, & \text{if } l_G^v(x,y)\cap \to i, \\ \pi_z(L)-1, & \text{if } i \in l_G^v(x,y),\ \pi_i(L) \neq 0, \\ \pi_z(L) & \text{in all other cases.} \end{cases} \tag{3}$$

Since the correctness of a prognosis cannot be verified immediately after it has been obtained, a correct prognosis issued by the digital computer must possess a property inherent in all correct prognoses: the initial sequence of states contained in the query must be included in the prognosis.

It may happen that a prognosis correct in this sense will not be confirmed by subsequent clinical observations. To some extent this is compensated by the fact that the physician can intervene in the learning process.

Let $v_i = 1$ and let the graph G contain the arcs ai and aj. Then the probability of obtaining a correct prognosis $\mathbf{w}$ when the relation (1) is used during training is equal to

$$P\{\mathbf{v} \subset \mathbf{w}\} = \prod_{i,\, v_i=1} \Phi\left(\frac{Mb_{ai}^v - Mb_{aj}^v}{\sqrt{Db_{ai}^v + Db_{aj}^v}}\right) P\{a \in l_G^v(x,y)\},$$

where $\Phi(x)$ is the density of the normal distribution law; Mb_{ai}^v is the mean value of the random variable b_{ai}; Db_{ai}^v is its variance.

If relations (2) or (3) are used in training, then

$$P\{\mathbf{v} \subset \mathbf{w}\} = \prod_{i,\, v_i=1} \Phi\left(\frac{p_{ai} g_i^v(N) - p_{aj} g_j^v(N)}{\sqrt{(p_{ai} g_i^v + p_{aj} g_j^v)(1 - p_{ai} g_i^v + p_{aj} g_j^v)}}\right) P\{a \in l_G^v(x, y)\},$$

where $g_i^v(N)$ is the probability that vertex i belongs to T_v after training with the aid of an o.p. whose length is N.

In carrying out experiments with a digital computer, the course of the operation and the postoperative period was considered for 60 patients operated on for Fallot' s tetralogy in the thoracic surgery clinic of the Kiev Scientific Research Institute of Tuberculosis and Thoracic Surgery named after F. G. Yanovsky. In the examination mode, 60 queries were proposed, which did not belong to any o.p. After training with the aid of an o.p. of total length 50, 57 prognoses were obtained. Of these, 52 were correct. Fifty prognoses were confirmed by subsequent observations in the clinic.

Institute of Cybernetics
Academy of Sciences of the Ukrainian SSR

Received
24 XII 1968

CITED LITERATURE

1. F. B. Gulko, B. Ya. Kogan, M. E. Raiskina, *Avtomatika i telemekhanika*, **8**, 104 (1967).

2. E. V. Gubler et al., in: *Cybernetics and Electronic Computing Technology*, 1962, p. 130.

3. V. M. Glushkov, *Introduction to Cybernetics*, 1963.

4. S. Ya. Zalavsky, in: *Biological Cybernetics*, **1**, 1969.

Note: Figure translations are in progress. See original paper for figures.