ON ONE CLASS OF LEARNING RECOGNITION ALGORITHMS

G. M. ADELSON-VELSKII, P. E. KUNIN, A. A. LEMAN

1967

SovietRxiv

Abstract

Full Text

UDC 519.95

CYBERNETICS AND THE THEORY OF REGULATION

G. M. ADELSON-VELSKII, P. E. KUNIN, A. A. LEMAN

ON ONE CLASS OF LEARNING RECOGNITION ALGORITHMS

(Presented by Academician M. V. Keldysh on 29 VII 1966)

Let A and B be disjoint subsets of a set C. An algorithm f is called a **recognition algorithm** for the sets A and B if, for every element $X \in A \cup B$, it either gives the answer that $X \in A$ (or $f(X) = A$), or that $X \in B$ (or $f(X) = B$), or refuses to answer ($f(X) = 0$).

In this paper a class of learning algorithms is described which, from given finite subsets $\bar{A} \subset A$ and $\bar{B} \subset B$, construct a recognition algorithm for the sets A and B.

In what follows it is assumed that C is the set of vertices $X(x_1, x_2, \dots, x_n)$ of the n-dimensional unit cube; the coordinates $x_1, x_2, \dots, x_n$ are called **features**.

M. M. Bongard ([1]) proposed finding combinations of features and their values that are characteristic of $\bar{A}$ and $\bar{B}$. The proposed iterative algorithms make it possible, in finding such combinations, to avoid complete enumeration and, consequently, to find combinations of a large number of features.

Definition 1. The **distance** between points $X \in C$ and $Y \in C$ in the system of features $(i_1, i_2, \dots, i_k)$ is called

$$\rho_{i_1, i_2, \dots, i_k}(X, Y) = \sum_{l=1}^{k} |x_{i_l} - y_{i_l}|. \tag{1}$$

Definition 2. A **tube** $T\{(i_1, i_2, \dots, i_k); X^0; r\}$ is the set of points $X \in C$ for which

$$\rho_{i_1, i_2, \dots, i_k}(X, X^0) \leq r. \tag{2}$$

The features $x_{i_1}, x_{i_2}, \dots, x_{i_k}$ are called **essential** for the tube T, X^0 is called the **center**, and r the **radius** of the tube. Obviously, the center of the tube T is any point of the hyperplane $\{x_{i_l} = x^0_{i_l}\}$, $(l = 1, 2, \dots, k)$.

Definition 3. A tube T is called **q-distinguishing** for the sets M and N and the function $\varphi(x)$, if

$$\frac{\nu[T \cap M \cap \Phi_-] + \nu[T \cap N \cap \Phi_+] + \nu[T \cap (M \cup N) \cap \Phi_0]}{\nu[T \cap (M \cup N)]} < q, \tag{3}$$

where $\nu[\Delta]$ is the number of elements of the set Δ; Φ_0, Φ_-, and Φ_+ are, respectively, the sets of points $X \in C$ for which $\varphi(X) = 0$, $\varphi(X) < 0$, $\varphi(X) > 0$. A special case of q-distinguishing tubes are q-distinguishing pure M-tubes ${}^{M}T$, for which $\varphi(X) = 1$, and pure N-tubes ${}^{N}T$, for which $\varphi(X) = -1$.

Definition 4. A system of tubes $\{T_1, T_2, \dots, T_s\}$ is called **complete** for a set D, if $D \subset \bigcup_i T_i$.

Let there exist for the set $A \cup B$ a complete system of q-distinguishing tubes $\{T_1, T_2, \dots, T_s\}$, where $s \ll \nu[A \cup B]$.

Obviously, for any subsets $\bar{A} \subset A$ and $\bar{B} \subset B$ there also exists a complete system of q-distinguishing tubes.

The algorithm for constructing q-distinguishing pure $\bar{A}$-tubes is as follows.

Let $T\{(i_1, i_2, \dots, i_k); X^0; R\}$ be some tube, $0 \le \delta_0 \le \delta_1 \le 1$,

$$\sigma_{\bar{A},l} = \sum_{X \in \bar{A} \cap T} x_l / \nu[\bar{A} \cap T], \qquad l = 1, 2, \dots, n. \tag{4}$$

The feature x_l is declared essential for the tube $\widetilde{T}_r\{(\tilde{i}_1, \tilde{i}_2, \dots, \tilde{i}_k); \widetilde{X}^0; r\}$ if $\sigma_{\bar{A},l} \le \delta_0$ or $\sigma_{\bar{A},l} \ge \delta_1$; in the first case $\tilde{x}^0_l = 0$, in the second $\tilde{x}^0_l = 1$.

Let

$$\Psi(\widetilde{T}_r) = \Psi(\nu[\bar{A} \cap \widetilde{T}_r],\, \nu[\bar{B} \cap \widetilde{T}_r]), \tag{5}$$

where $\Psi(n_1, n_2)$ is a given function, monotonically increasing in the first variable and monotonically decreasing in the second. The radius $\widetilde{R}$ of the tube $\widetilde{T}\{(\tilde{i}_1, \tilde{i}_2, \dots, \tilde{i}_k); \widetilde{X}^0; \widetilde{R}\}$ is chosen so that the value $\Psi(\widetilde{T}_R)$ is maximal among all $\Psi(\widetilde{T}_r)$ for $\nu[\bar{A} \cap \widetilde{T}_r] > \gamma$. Thus, the algorithm is iterative.

As the initial center one may choose an arbitrary point $X \in \bar{A}$, and as the essential features of the initial tube, all features $x_1, x_2, \dots, x_n$. The number of iterations may be specified in advance; one may also continue the iterations as long as the quality of the tubes obtained does not deteriorate. For the last iteration, in addition, it is required that $\widetilde{T}$ be a q-distinguishing pure $\bar{A}$-tube, i.e. $\Psi(\widetilde{T}) < q$.

The process described above may also fail to lead to the construction of a q-distinguishing pure tube; in that case, a new point $X \in \bar{A}$ must be chosen as the initial one.

The bounds δ_0 and δ_1 may depend on the feature number l. In one variant of the algorithm these bounds are determined by the formulas

$$\delta_{0,l} = \sum_{X \in \bar{B}} x_l / \nu[\bar{B}] - \Delta_0, \tag{6}$$

$$\delta_{1,l} = \sum_{X \in \bar{B}} x_l / \nu[\bar{B}] + \Delta_1, \tag{6'}$$

where the standards Δ_0, Δ_1 are given.

Suppose that a system of q-distinguishing pure $\bar{A}$-tubes $T_1, \ {}^{A}T_2, \dots, \ {}^{\bar{A}}T_\alpha$ has already been constructed. If it is not complete for the set $\bar{A}$ and not all points $X \in \bar{A}$ have been tried as initial centers, then the process of constructing new q-distinguishing pure $\bar{A}$-tubes can be continued. As the initial center one chooses, for example, a point $X \in \bar{A}$ that is farthest from all already constructed centers (in the metrics of the corresponding tubes). A system of q-distinguishing pure $\bar{B}$-tubes is constructed analogously.

After a system of q-distinguishing pure A- and B-tubes has been obtained, the algorithm f is constructed so that

$$f(X) = A, \quad \text{if } X \in \bigcup_i {}_{\bar{A}}T_i \setminus \bigcup_j {}_{\bar{B}}T_j,$$

$$f(X) = B, \quad \text{if } X \in \bigcup_j {}_{\bar{B}}T_j \setminus \bigcup_i {}_{\bar{A}}T_i,$$

$$f(X) = 0 \quad \text{in all other cases.}$$

The variant of the algorithms of this class described above was implemented as a computer program. Testing of the algorithm showed that it successfully finds pure A- and B-tubes for which the probability of is sufficiently large ($c > 1/\sqrt{m}$). At the same time, when the number of elements is small

The $\bar{A} \cup \bar{B}$ algorithm has a tendency to create "prejudices," i.e., to seek pure $\bar{A}$- and $\bar{B}$-tubes that are not A- and B-tubes.

The process of constructing impure q-distinguishing tubes is also iterative. The function $\varphi(X)$ is defined as follows. Let $T\{(i_1, i_2, \dots, i); X^0; r\}$ be a tube; then

$$\varphi(X) = \omega_{\bar{A}}(X) / \{[\omega_{\bar{A}}(X) + \omega_{\bar{B}}(X)] - 1/2\},$$

where

$$\omega_{\Delta}(X) = \nu[\Delta \cap T] \prod_i \frac{\nu\{X' \in \Delta \cap T : x_i' = x_i\}}{\nu[\Delta \cap T]},$$

the product being taken over all features that are inessential for the tube T. These formulas follow from the assumption that the features inessential for the tube T are uncorrelated for the elements of the tube, and that the mean values of these features are close to the probabilities $P\{x_i = 1\}$.

Institute
of Theoretical and Experimental Physics

Received
18 VI 1966

CITED LITERATURE

1. M. M. Bongard, *Biofizika*, **6**, No. 2 (1961).

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