Kybernetika

Alexander A. Georgiev

The strong pointwise convergence of nearest neighbor function fitting algorithm with applications to system identification

Kybernetika, Vol. 21 (1985), No. 2, 118--122

Persistent URL: http://dml.cz/dmlcz/124802

KYBERNETIKA — VOLUME 21 (1985), NUMBER 2

THE STRONG POINTWISE CONVERGENCE OF NEAREST NEIGHBOR FUNCTION FITTING ALGORITHM WITH APPLICATIONS TO SYSTEM IDENTIFICATION

ALEXANDER A. GEORGIEV

A new nonparametric algorithm, based on the nearest neighbor rule, is investigated in the case when noisy measurements of real valued function g are taken in nonrandom domain points. Sufficient conditions for the strong pointwise convergence of the procedures are given. Applications for a wide class of system identification problems are discussed.

1. INTRODUCTION

An important problem in system and control engineering is the identification of a system g from the observations

$$\{x_i, g(x_i) + Z_i\}_{i=1}^{n}, \tag{1}$$

where x_i is the input of the system, selected by the experimenter, and Z_i is an independent zero-mean noise. If the system g is known except for a finite number of parameters, the most popular methods are the least square [5] or maximum likelihood methods [5], [9]. If g is completely unknown, the following model may be proposed

$$\tilde{g}(x) = \boldsymbol{a}^{\mathrm{T}} \boldsymbol{\varphi}(x),$$

where $\boldsymbol{\varphi}^{\mathrm{T}}(x) = [\varphi_1(x), \ldots, \varphi_l(x)]$ are known basis functions and $\boldsymbol{a}$ is a vector of parameters which is estimated by different methods such as stochastic approximation [15] or random search algorithm [16].

In this paper it is assumed that the system g is unknown and moreover, no parametric models are assumed. Based on the data sequence (1) we construct a new nearest neighbor identification algorithm. In Section 2 of this paper we prove that the algorithm is strongly pointwise consistent. We also show that the procedure is applicable to the identification of the step and frequency response in linear dynamic continuous-time systems of an unknown order.

2. THE ALGORITHM AND ITS CONVERGENCE

Let (x_1, Y_1), $(x_2, Y_2), \ldots, (x_n, Y_n)$ be observations according to the nonlinear memoryless system

$$Y_i = g(x_i) + Z_i , \quad i = 1, 2, \ldots, n ,$$

where g is an unknown function defined on the closed interval $[0, 1]$ and the errors Z_i are independent, identically distributed random variables with zero mean and finite variance. The input signals $\{x_i\}$ are assumed to be known without errors and they satisfy, without loss of generality, the order condition

$$x_1 < x_2 < \ldots < x_n .$$

Let $x_0 = 0$ and $x_{n+1} = 1$ and let $r_n = r(x, k_n)$ be the distance between x and the k_nth nearest neighbor observation of x among all x_i. The number of neighbors k_n depends on n and is determined by the experimenter. In this paper we propose the following new nearest neighbor algorithm for the estimation of g

$$g_n(x) = \sum_{i=1}^{n} Y_i K\left(\frac{x - x_i}{r_n}\right) \frac{(x_i - x_{i-1})}{r_n} , \tag{2}$$

where K is a bounded nonnegative function defined on the real line. Another kernel procedure has been studied by Priestley and Chao [10], Benedetti [1] and Schuster and Yakowitz [17]. Rutkowski [13] has introduced an orthogonal series estimate for g. Another nearest neighbor procedure has been examined by Greblicki [7] for similar identification problems.

The following theorem establishes sufficient conditions for strong consistency of the algorithm (2).

Theorem. Let g be a bounded function, and K be a continuous probability density function, such that $K(x)$ is nonincreasing for $x > 0$, and nondecreasing for $x < 0$. Let

$$\Delta_n = \max_i (x_i - x_{i-1}) \leqq \beta/n , \quad \beta > 0 , \tag{3}$$

and

$$\delta_n = \min_i (x_i - x_{i-1}) \geqq \alpha/n , \quad \alpha > 0 . \tag{4}$$

If

$$k_n/n \to 0 , \quad \text{as} \quad n \to \infty , \tag{5}$$

and

$$k_n^{-1} = O(n^{-t}) , \tag{6}$$

$$\mathsf{E}Z_1 = 0 , \quad \mathsf{E}|Z_1|^{1+1/t} < \infty , \quad \text{for some } 0 < t < 1, \tag{7}$$

then

$$g_n(x) \to g(x) \quad \text{with probability 1, as} \quad n \to \infty , \tag{8}$$

at every continuity point of $g(x)$.

Proof. Let $I_{\{|x| \leqq a\}}$ be the indicator of the set of all points for which $|x| \leqq a$. By definition of r_n, we obtain the following relation between k_n and r_n

(9) $$k_n \delta_n \leqq r_n \leqq k_n \Delta_n .$$

Let us choose $a > 0$ and x as continuity point of g. It is obvious that

$$|g_n(x) - g(x)| \leqq |g_n(x) - \mathsf{E}g_n(x)| + |\mathsf{E}g_n(x) - g(x)| .$$

For the bias of estimate (2) we can give the following upper bound

(10) $$|\mathsf{E}g_n(x) - g(x)| \leqq$$
$$\leqq \sum_{i=1}^{n} K\left(\frac{x - x_i}{r_n}\right) \frac{(x_i - x_{i-1})}{r_n} |g(x_i) - g(x)| I_{\{|x_i - x| \leqq a\}} +$$
$$+ \sum_{i=1}^{n} K\left(\frac{x - x_i}{r_n}\right) \frac{(x_i - x_{i-1})}{r_n} |g(x_i) - g(x)| I_{\{|x_i - x| > a\}} +$$
$$+ \left|\sum_{i=1}^{n} K\left(\frac{x - x_i}{r_n}\right) \frac{(x_i - x_{i-1})}{r_n} - 1\right| |g(x)| \leqq$$
$$\leqq \varepsilon b + 2c \sum_{i=1}^{n} K\left(\frac{x - x_i}{r_n}\right) \frac{(x_i - x_{i-1})}{r_n} I_{\{|x_i - x| > a\}} + c\left|\sum_{i=1}^{n} d_{in} - 1\right| ,$$

where $c = \sup_{x \in [0,1]} |g(x)|$, $\varepsilon = \sup_{|y - x| \leqq a} |g(y) - g(x)|$, $d_{in} = K\left(\frac{x - x_i}{r_n}\right) \frac{(x_i - x_{i-1})}{r_n}$ and $\sum_{i=1}^{n} d_{in} \leqq b$. Let first $n \to \infty$. From Lemma 1 in Benedetti [1], p. 250, and by assumptions from the Theorem, we get for the third term of (10) that

(11) $$\sum_{i=1}^{n} d_{in} \to 1 , \quad \text{as} \quad n \to \infty .$$

Obviously $b < \infty$, because $\{d_{in}\}$ is a sequence of nonnegative numbers and (11) is in force. The second term in (10) is not greater than

$$2c \sum_{i=1}^{n} K\left(\frac{x - x_i}{r_n}\right) \frac{|x - x_i|}{|x - x_i|} \frac{(x_i - x_{i-1})}{r_n} I_{\{|x - x_i| > a\}} \leqq \frac{2cn\Delta_n}{a} \sup_{|y| > a/r_n} K(y)|y| .$$

Now we use (3), (5) and (9). We obtain that (10) convergences to zero if we first let $n \to \infty$ and then $a \to 0$. It is enough now to show that $|g_n(x) - \mathsf{E}g_n(x)|$ tends to zero *with probability* 1, as $n \to \infty$. Obviously

$$|g_n(x) - \mathsf{E}g_n(x)| = \left|\sum_{i=1}^{n} d_{in} Z_i\right| .$$

It follows, from (3), (4) and (9), that

$$\max_i |d_{in}| \leqq \frac{\beta \sup K(x)}{\alpha} \frac{1}{k_n} .$$

By virtue of Pruitt's result [11], p. 769, and (6), (7) the proof is complete. □

3. APPLICATIONS IN DYNAMICAL SYSTEM IDENTIFICATION

The study of procedure (2) was motivated in part by our interest in identification of linear dynamical continuous-time systems using noisy measurements on a system output. Consider the dynamic system of an unknown order (see also Rutkowski [13]), with zero initial conditions, given by

$$g(t) = \int_0^t w(t - s)\, u(s)\, \mathrm{d}s\,,$$

where $w(t)$ is the weight function of the system, $u(t)$ is the input signal $(u(t) = 0$ for $t < 0)$ and $g(t)$ is the output signal. The system is observed as

$$Y(t) = g(t) + Z(t)\,, \tag{12}$$

where $Z(t)$ is the measurement noise with zero mean, finite variance and such that $Z(t')$ and $Z('')$ are independent, identically distributed random variables for $t' \neq t''$. For the unit-step input the relation (12) becomes

$$Y(t) = h(t) + Z(t)\,,$$

where $h(t)$ is the step response. Let us suppose that n measurements are to be made and are carried out at equally spaced points $1/n, 2/n, \ldots, 1$ in the unit interval. Thus, the noisy values $\{Y_i\}_{i=1}^n$ are recorded, where Y_i's are independent, and $h(t)$ may be estimated by algorithm (2).

Procedure (2) can also be used for identification the frequency response of a linear dynamic system (see Greblicki [7]). Assume that $H(\mathrm{j}\omega)$ is a transfer function and the input signal is $u(t) = \sin(\omega t)$. It is known that the output signal is

$$g(t) = a(\omega) \sin(\omega t + \varphi(\omega))\,,$$

where $a(\omega) = |H(\mathrm{j}\omega)|$ and $\varphi(\omega) = \arg H(\mathrm{j}\omega)$. Let us suppose that $2n$ noisy measurements are to be made in $\{\omega_i\}_{i=1}^n$, i.e.

$$A_i = |H(\mathrm{j}\omega_i)| + Z_{1i}\,, \quad i = 1, 2, \ldots, n\,,$$

and

$$\Phi_i = \arg H(\mathrm{j}\omega_i) + Z_{2i}\,, \quad i = 1, 2, \ldots, n\,.$$

The noisy measurements $\{A_i\}_{i=1}^n$ and $\{\Phi_i\}_{i=1}^n$ and frequency points $\{\omega_i\}_{i=1}^n$ are enough for consistent estimating $a(\omega)$ and $\varphi(\omega)$ by algorithm (2).

For an other nonparametric procedures in system identification, the reader is referred to Greblicki [6], [7], Greblicki and Krzyżak [8], Rutkowski [13], [14], Georgiev [2], [3], [4], as well as Rafajłowicz [12].

ACKNOWLEDGMENT

This work was done as part of my Ph. D. dissertation under the supervision of Professor Włodzimierz Greblicki, at the Technical University of Wrocław; I am grateful to him for his many very helpful suggestions.

(Received April 15, 1983.)

REFERENCES

[1] J. Benedetti: On the nonparametric estimation of regression function. J. Roy. Statist. Soc. *B 39* (1977), 248–253.

[2] A. A. Georgiev: Nonparametric system identification by kernel methods. IEEE Trans. Automat. Control *AC-29* (1984), 356–358.

[3] A. A. Georgiev: A nonparametric algorithm for identification of linear dynamic SISO systems of unknown order. Systems & Control Letters *4* (1984), 273–280.

[4] A. A. Georgiev: Nonparametric kernel algorithm for recovery of functions from noisy measurements with applications. IEEE Trans. Automat. Control *AC-30* (1985), to appear.

[5] G. C. Goodwin and P. L. Payne: Dynamical System Identification: Experiment Design and Data Analysis. Academic Press, New York 1977.

[6] W. Greblicki: Nonparametric system identification by orthogonal series. Problems Control Inform. Theory *8* (1979), 67–73.

[7] W. Greblicki: Nearest Neighbor Algorithm for Recovering Functions from Noise. Report No. 30/82/I-6, Technical University of Wrocław, Wrocław 1982.

[8] W. Greblicki and A. Krzyżak: Nonparametric identification of a memoryless system with cascade structure. Internat. J. System Sci. *10* (1979), 1301–1310.

[9] R. L. Kashyap: Maximum likelihood identification of stochastic linear systems. IEEE Trans. Automat. Control *AC-15* (1970), 25–34.

[10] M. B. Priestley and M. T. Chao: Nonparametric function fitting. J. Roy. Statist. Soc. *B 34* (1972), 385–392.

[11] W. E. Pruitt: Summability of independent random variables. J. Math. Mech. *15* (1966), 769–776.

[12] E. Rafajłowicz: Nonparametric algorithm for identification of weakly nonlinear static distributed-parameter systems. Systems & Control Letters *4* (1984), 91–96.

[13] L. Rutkowski: On system identification by nonparametric function fitting. IEEE Trans. Automat. Control. *AC-27* (1982), 225–227.

[14] L. Rutkowski: On nonparametric identification with prediction of time-varying systems. IEEE Trans. Automat. Control *AC-29* (1984), 58–60.

[15] G. N. Saridis, Z. J. Nicolic and K. S. Fu: Stochastic approximation for system identification, estimation, and decomposition of mixtures. IEEE Trans. Syst. Sci. Cybern. *SSC-5* (1969), 8–15.

[16] G. N. Saridis: Expanding subinterval random search for system identification and control. IEEE Trans. Automat. Control. *AC-22* (1977), 405–412.

[17] E. F. Schuster and S. Yakowitz: Contributions to the theory of nonparametric regression with application to system identification. Ann. Statist. *7* (1979), 139–149.

Dr. Alexander A. Georgiev, Institute of Engineering Cybernetics, Technical University of Wrocław, Wyspianskiego 27, I-6, 50-370 Wrocław. Poland.