neural network training,
conclusion-making systems,
medical diagnostics automation

Sławomir CHANDZLIK*
Jan PIECHA *, **

A PATIENT WALK-DATA-RECORD MODELLING USING A SPLINE INTERPOLATION METHOD

The paper describes an analytical method of data record description that allows converting samples of discrete data record into continuous function. This operation allows re-sampling the data record with a sampling rate that is adequate to step duration. The record length is limited to an efficient size for training the Conclusion-Making Unit (CMU). Various options available in the PSW equipment [6], [7] give the user many aims in putting diagnosis anyhow, due to simplification of the CMU training process several methods for data records modifications are considered.

1. INTRODUCTION

The data record structure, being a subject of analysis, has been described in many papers [6], [7], [2]. The Parotec System for Windows (PSW) is a pedobarograph (data recorder) used for data collection. The pressure load spectrum registered by set of sensors (on insole) provides the user with interfaces describing a foot shape and way of walking [1], [10]. For this analysis several two- and three-dimensional interfaces were made. They are used for diagnosis support in orthopaedics and neurology.

The neural network unit has been offered as an automatic conclusion-making-unit (CMU) [2]. The effectiveness of the CMU training depends not only on the network topology structure or complexity [12], [5] but also on length of the data record used for training procedure.

That is why many affords have been undertaken for the data record length minimisation [12], [5]. The data reader samples the pressure values on each sensor of the insole. The PSW data record consists of discrete pressure values recorded on every sensor. The 24 pressure values are measured with sampling intervals defined by micro-controller internal clock. Although operator, in range of 100 to 200 Hz can select the sampling frequency of the data recorder, a random walking speed produces data records with different lengths [3].

Various sampling frequencies are explainable for distinguishing running from walking activities. Anyhow, statistically variable speed of a patient walk, with a constant clock period of the

* University of Silesia, Institute of Informatics, Dept. Computer Systems, Katowice, Poland.
** Silesian University of Technology, Institute of Transport, Katowice, Poland.

data controller produces different data records. The same disease is represented by different data sequences.

The paper indicates several problems that have to be solved for normalisation of the data record used for neural network (CMU) training.

One of the normalisation methods uses polynomial interpolations, described in the paper [3]. This method causes many errors that can be avoided by speculations with polynomials levels. Anyhow many troubles with calculations stability were observed (The Runge phenomena) [3]. These additions to the approximation algorithm mean that the calculation complexity is growing remarkable.

The spline method of the data samples approximation simplifies many steps in normalisation algorithms. What is more there are none problems with the calculations stability.

2. THE PRESSURE DISTRIBUTION APPROXIMATION USING A SPLINE FUNCTIONS

In computer algorithms the spline functions for discrete sequences approximation are often used [4], [8]. They are very easy to define and the interpolation functions are concurrent to functions that are a subject of interpolation process when the interpolation ranges are small enough [11].

For many cases interpolation functions are smooth and concurrent [4]. According to literature recommendations [4], [11], [9] the spline functions of third level have been applied. For this functions continuity of a second differential coefficient is a measure of smoothness.

Let us assume that all data vectors are defined the same way as it was defined in work [3]. It means that the input vector consists of:

$$l_p = 24 * 50 = 1200 \text{ components,} \tag{1}$$

Each component contains 50 samples creating well-ordered set of values:

$$X = \{x_i : i = 1,2,...,50\} \tag{2}$$

where: $x_1 < x_2 < ... < x_{50}$ are values of input variables that define output set Y_j - discrete values of pressure in each interval i for every sensor j:

$$Y_j = \{f_i : i = 1,2,...,50\} \quad j = 1,2,...,24\,. \tag{3}$$

The transform of normalisation is defined by following relations:

$$\begin{aligned} &F_j : R_X \to R^+ \cup \{0\} \quad R_X = \{x : x \in R \wedge x_1 \le x \le x_{50}\};\, x_1, x_{50} \in X \\ &F_j(x_i) = f_i \qquad\qquad x_i \in X;\; f_i \in Y_j;\; j = 1,2,...,24 \end{aligned} \tag{4}$$

The $s(x)$ in interval $[a, b]$ is the spline function of third level, if:

1. $s(x)$ is a polynomial of no higher then third level, in every sub-interval, $(x_i, x_{i+1}) \subset [a, b]$ (i=1, 2, …, n+1),
2. $s(x) \in C^2([a, b])$.

A set of all third level spline functions for nodes defined by variables x_i has been assigned $S_3(\Delta_n)$, where: Δ_n is an interval $[a, b]$ division for n sub-intervals. Then additional assumptions concern:

1. The interpolation condition (4): $s(x_i) = f_i$; $s(x_{i+1}) = f_{i+1}$ (i=1, 2, …, n);
2. The continuity condition: $s^{(k)}(x_i)$ continue (i = 2, 3, …, n; k = 1, 2);
3. The natural splines: $s''(x_1) = s''(x_n) = 0$;

The above assumptions and descriptions allow to define spline transformation (4) functions $s(x)$:

$$s(x) = f_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3 \quad \textit{for } x \in [x_i, x_{i+1}], \tag{5}$$

where: constant values of b_i, c_i, d_i are defined for every interval, from equations:

$$\begin{aligned} b_i &= \frac{y_{i+1} - y_i}{x_{i+1} - x_i} - (x_{i+1} - x_i)(\sigma_{i+1} + 2\sigma_i) \\ c_i &= 3\sigma_i \\ d_i &= \frac{\sigma_{i+1} - \sigma_i}{x_{i+1} - x_i} \\ \sigma_i &= s''(x_i) \end{aligned} \tag{6}$$

for i = 1, 2, …, n-1.

This way the spline functions were used for pressure distribution on 24 sensors of each insole. Four groups of data patterns have been selected for experiments, as: the group of comparison, one-sided ischialgia, bunion pathology and paresis. For each group the pattern record has been selected.

3. AN ERROR ANALYSIS OF INTERPOLATION BY SPLINE FUNCTIONS

The effectiveness of interpolation methods was carried out on error analysis. In each classical error analysis using spline functions we assume continuity of derivatives of functions considered for interpolation principles according to the assumption ($f \in C^4([a,b])$) [4], [8], [11]).

As in the presented subject the above condition is not available this criterion can not be used. That is why the absolute and relative errors-analysis has been made, for data record values in sampling nodes only. According to the literature [3] the faults expressions are defined in a following manner:

The absolute error:
$$E_i = \left| f_i - F(x_i) \right| \qquad i = 1,2,...,50 \tag{7}$$

The relative error:
$$RE_i = \begin{cases} \dfrac{E_i}{f_i} & when\ f_i \neq 0 \\ E_i & when\ f_i = 0 \end{cases} \qquad i = 1,2,...,50 \tag{8}$$

where: f_i – concerns values in points $x_i \in X$,
$F(x_i)$ – is the approximation result (for each $x_i \in X$).

Faults for single sensors are defined by following formulas:

Average absolute error:
$$E_j = \frac{\sum_{i=1}^{50} E_i}{50} \tag{9}$$

Average relative error:
$$RE_j = \frac{\sum_{i=1}^{50} RE_i}{50} \tag{10}$$

Maximal absolute error:
$$E_{\max j} = \max_i \{E_i : i = 1,2,...,50\}, \tag{11}$$

Maximal relative error:
$$RE_{\max j} = \max_i \{RE_i : i = 1,2,...,50\}, \tag{12}$$

where: j = 1, 2, …, 24.

The interpolation procedure starts from the data record division into n = 49 samples, according to the sampling rate for data record [1, 50] samples. For every interval the spline function of third level has been defined.

The approximations obtained by formulas of 7 ÷ 12 were defined with not remarkable errors of calculations <= 5.0E–4. When the division level is smaller the approximation error is a bit higher. For the division level n = 12 the approximation error is equal to errors obtained for polynomial interpolation algorithms [3] that are presented in table 1.

sensor no	aver. relative err.[1]			aver. absolute err.[2]			max. relative err.[3]			max. absolute err.[4]		
	p49[5]	p12[6]	iwp[7]	p49	p12	Iwp	p49	p12	iwp	p49	p12	iwp
1	0,000	0,030	0,076	0,000	0,435	0,326	0,000	0,500	3,004	0,000	7,201	5,142
2	0,000	0,088	0,149	0,000	0,465	0,323	0,000	0,857	2,072	0,000	5,654	3,310
3	0,000	0,030	0,030	0,000	0,171	0,152	0,000	0,476	0,487	0,000	1,499	1,274
4	0,000	0,031	0,048	0,000	0,151	0,171	0,000	0,286	0,602	0,000	0,527	0,678
5	0,000	0,035	0,049	0,000	0,074	0,073	0,000	0,753	1,354	0,000	0,425	0,468
6	0,000	0,126	0,225	0,000	0,031	0,052	0,000	0,425	0,974	0,000	0,142	0,250
7	0,000	0,045	0,034	0,000	0,074	0,061	0,000	0,677	0,828	0,000	0,436	0,310
8	0,000	0,296	0,136	0,000	0,043	0,050	0,000	3,868	1,664	0,000	0,176	0,165
9	0,000	0,072	0,087	0,000	0,007	0,005	0,000	0,602	0,909	0,000	0,071	0,038
10	0,000	0,137	0,048	0,000	0,086	0,084	0,000	3,172	0,664	0,000	0,472	0,441
11	0,000	0,108	0,068	0,000	0,035	0,034	0,000	1,402	0,411	0,000	0,155	0,116
12	0,000	0,075	0,062	0,000	0,018	0,012	0,000	0,536	0,321	0,000	0,158	0,075
13	0,000	0,117	0,058	0,000	0,145	0,146	0,000	2,085	0,851	0,000	0,981	1,013
14	0,000	0,104	0,052	0,000	0,108	0,122	0,000	2,148	1,016	0,000	0,824	0,475
15	0,000	0,065	0,043	0,000	0,125	0,067	0,000	0,929	0,944	0,000	1,325	0,324
16	0,000	0,120	0,074	0,000	0,109	0,112	0,000	3,501	1,899	0,000	1,559	1,380
17	0,000	0,038	0,053	0,000	0,103	0,098	0,000	0,834	1,336	0,000	0,862	0,435
18	0,000	0,125	0,016	0,000	0,400	0,163	0,000	4,084	0,211	0,000	4,287	1,392
19	0,000	0,099	0,028	0,000	0,362	0,237	0,000	2,465	0,437	0,000	4,470	2,410
20	0,000	0,104	0,049	0,000	0,262	0,261	0,000	2,812	1,006	0,000	2,320	1,380
21	0,000	0,069	0,065	0,000	0,062	0,064	0,000	1,682	2,050	0,000	0,333	0,336
22	0,000	0,049	0,036	0,000	0,129	0,105	0,000	0,359	0,285	0,000	0,689	0,504
23	0,000	0,037	0,050	0,000	0,135	0,098	0,000	0,287	0,991	0,000	1,368	1,001
24	0,000	0,037	0,066	0,000	0,280	0,244	0,000	0,886	2,281	0,000	3,388	2,168
sum	0,0	2,0	1,6	0,0	3,8	3,1	0,0	35,6	26,6	0,0	39,3	25,1

Tab.1. Spline and Polynomial interpolation errors comparison

[1] average relative error
[2] average absolute error
[3] maximal relative error
[4] maximal absolute error
[5] spline for n=49
[6] spline for n=12
[7] polynomial interpolation results for algorithms defined in [3]

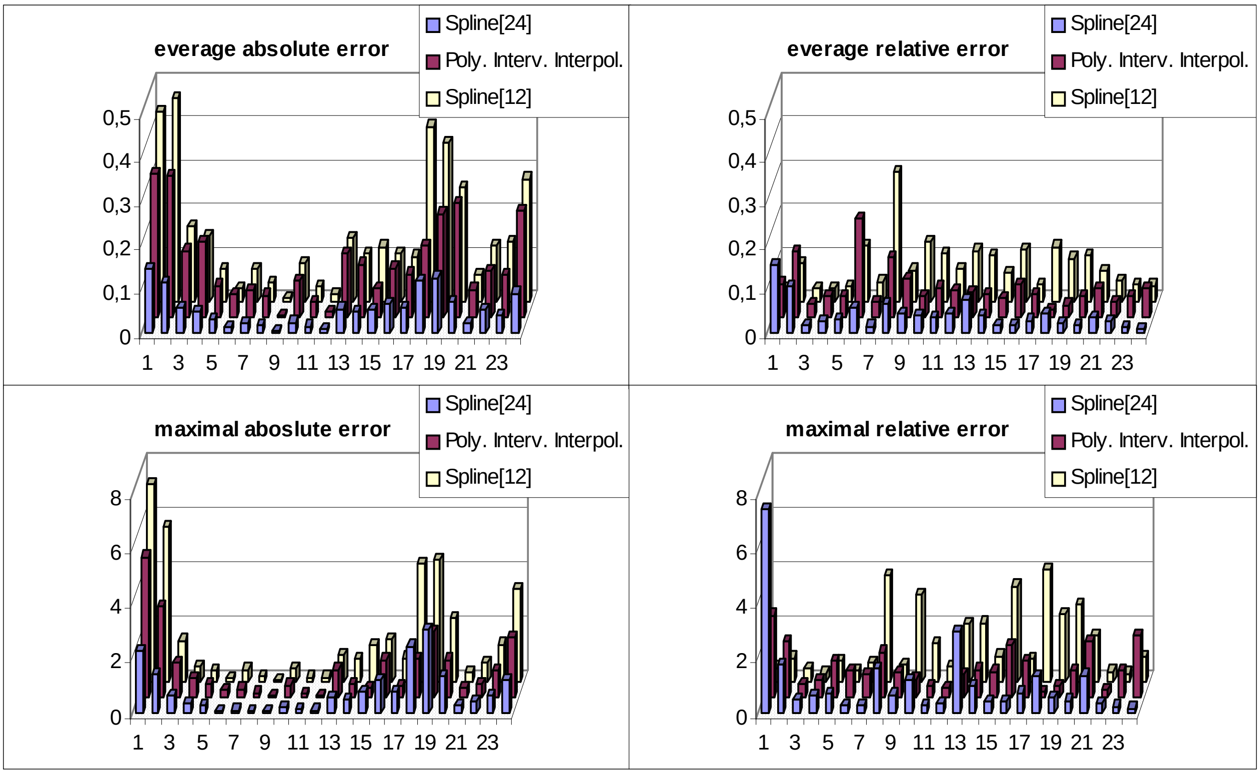

Fig.1. The error visualisation for spline of 24 (Spline[24]) and 12 (Spline[12]) intervals and for Polynomial Interval Interpolation (Pol. Interv. Interpol.)

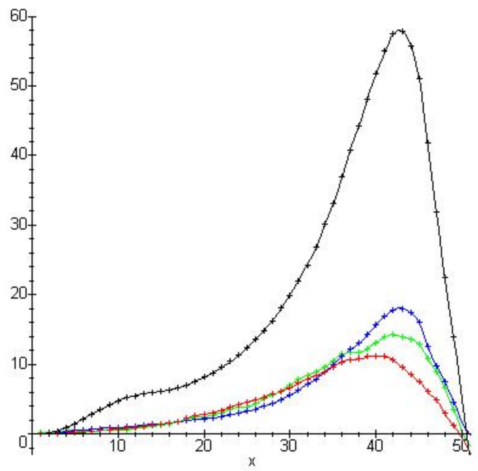

Fig.2. Spline interpolation at n=49 parameter.

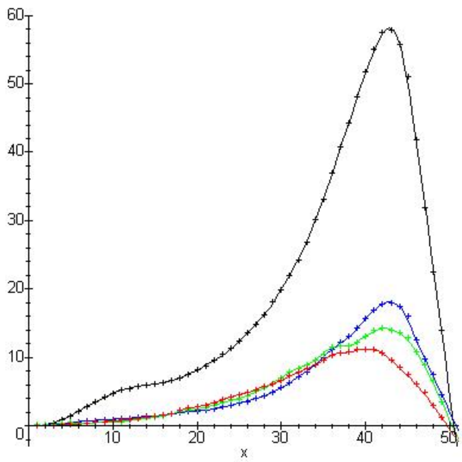

Fig.3. Spline interpolation at n=12 parameter.

4. CONCLUSIONS

Comparing the experiment results we can conclude that approximation algorithm using the spline functions are more effective then the polynomial interpolation methods [3], noticed as an classical approach. First of all the calculation complexity is reasonable smaller, with an error level of calculations not higher then the class of the data measuring and calculations quality. What is more the spline function method allows to reach the approximation errors close to zero, not using any tricky procedures that are used in the polynomial interpolation methods.

There are also several difficulties in spline functions approximation as values of second level differential coefficients definition (6). For this purpose a three-diagonal matrixes were successfully offered. According to literature [4] the operations complexity, with these matrixes, can be simplified to the level afforded by personal computer calculations power.

The CMU training depends on the input data quality. The presented algorithm helps reducing the time of CMU training process.

BIOBLIOGRAPHY

[1] ARITOMI H., MORITA M., YONEMOTO K., A simple method of measuring the foot sole pressure of normal subjects using prescale pressure-detection sheets. *J. Biomech. 16,* 1983, pp 157-165.

[2] CHANDZLIK S., KOPICERA K., Experiments with neural network parameters-selection for Foot abnormalities Recognition. Journal of Medical Informatics and Technologies, Vol. 5, pp. CS-71 – CS-78, November 2000.

[3] CHANDZLIK S., PIECHA J., Modelling the data record of a patient walk by Langrange-polynomial method. Journal of Medical Informatics and Technologies. Vol. 3/2002, pp. MI-143-152, Katowice 2002.

[4] FORTUNA Z., MACUKOW B., WĄSOWSKI J., Metody numeryczne. WNT, Warszawa 1982.

[5] OSOWSKI S., Sieci neuronowe w ujęciu algorytmicznym. WNT, Warszawa 1996 (in polish)

[6] PIECHA J., The neural network conclusion-making system for foot abnormality recognition. Proceedings of IMACS World Congress, Lausanne, Switzerland, August 2000.

[7] PIECHA J., The neural network selection for a medical diagnostic system using an Artificial data set. CIT JOURNAL, Vol. 9, No. 2, June 2001.

[8] PRESS W.H., TEUKOLSKY S.A., VETTERLING W.T., FLANNERY B.P., Numerical Recipes in C. The Art of Scientific Computing, Cambridge University Press, Melbourne 1992.

[9] RALSTON A., Wstęp do analizy numerycznej. PWN, Warszawa 1971.(in polish)

[10] SHEREF M.J., BREGMAN A.M., Kummer F.J., The effect of immobilisation devices on the load distribution under the foot. *Clinical Orthopedics 192*, 1985, pp. 260-267.

[11] STOER J., Wstęp do metod numerycznych. PWN, Warszawa 1979. (in polish)

[12] TADEUSIEWICZ R., Sieci neuronowe. Akademicka Oficyna Wydawnicza RM, Warszawa 1993. (in polish)