INFORMATICA, 2001, Vol. 12, No. 3, 431–438

Pre-processing of Data for Nonlinear Mapping

Algirdas Mykolas MONTVILAS
Institute of Mathematics and Informatics
A. Goštauto 12, 2600 Vilnius, Lithuania
e-mail: montvila@ktl.mii.lt

Received: May 2001

Abstract. The sequential nonlinear mapping is suitable for sequential detection of states of dynamic systems (Montvilas, 1999a). In addition, it can indicate the undesirable states and even the damages of dynamic systems. The last is complicated when the damage is caused by a small changing of respective parameter describing the state. In the paper the problem of nonlinear mapping to be sensitive for the small changing of parameters and the problem related with dimensionality of parameters are solved by using a special pre-processing of data. Examples are given.

Key words: nonlinear mapping, dynamic systems, states estimation, pre-processing of data.

1. Introduction

In the paper (Sammon, 1969) the method of *simultaneous* nonlinear mapping for data structure analysis is described. It allows aposteriori clustering, classification and visualisation of data. This method does not fit for the *sequential* processing of data. A triangulation method for the sequential mapping of points from N-space to two-space (Lee *et all*., 1977) preserves only two distances to vector previously mapped and, in addition, it uses the spanning tree, so it makes the mapping dependent on the history, hence its using is restricted. In the paper (Montvilas, 1995) the method of *sequential* nonlinear mapping for data structure analysis is considered. It allows to do that apriori (in real time). In addition, the *sequential* nonlinear mapping appeared to be suitable for sequential detection of states of dynamic systems (Montvilas, 1999a). The states of dynamic system can be unknown and can change themselves abruptly or slowly. The method is based on sequential nonlinear mapping of multi-dimensional vectors of parameters (collection of which describes the present state of dynamic systems) into two-dimensional vectors in order to reflect the states and their changes on the PS screen and to observe the situation by means of computer. The sequential nonlinear mapping can be applied for supervision and control of technological process and indication of its damage (Montvilas, 1999b).

In the paper we consider the problems related with relatively small changing of parameters and with variety of their dimensions. Some times parameter describing the state of a dynamic system has to be constant at every state of the dynamic system, and a small change of it governs a damage of the dynamic system. However the small changing of parameter is allowed for every of them, and the mapping does not indicate any damage.

The problem of making the sequential nonlinear mapping to be sensitive for small changing of parameters and the problem related with dimensions are solved by using a special pre-processing of data.

2. Problem Formulation

The essence of nonlinear mapping is to preserve the inner structure of distances among the parameter's vectors being in L-dimensional space (L – number of parameters) after mapping them to two-dimensional space (screen). The distances geometrically depend on parameter's numerical values and their changes. Some parameters describing the present state can be of great numerical value, other parameters can be of small one. According them the distances among the mapped vectors which represent on the screen the different states of the dynamic systems can be relatively large or small. This causes some troubles in identification of states of the dynamic system. Furthermore there are such situations when a small changing of one parameter causes the changing of state of the dynamic system or even indicates the damage of it. For example, the frequency of voltage in European power stations has to be 50Hz. The value of 49Hz or 51Hz indicates the trouble. Such power station has to be switched off power net. Concerning voltage a request is more easy. The value of 230V or 210V indicates only the undesirable state of the power station and call for improve that.

Let the states of imaginary power station (PS) be described by six parameters. The first parameter means a frequency, the second – a voltage. These two parameters have to be constant. However it is necessary to watch them because their change means either undesirable state or a damage of the power station. The remain four parameters mean some other important parameters of the imaginary power station's stable states. Let the power station has five stable working states and be watched at 25 time moments. The values of parameters are presented in Table 1.

The important time moments and values of parameters are in bold in the Table 1. The sequential nonlinear mapping requires the first M time moments parameter's vectors to be mapped simultaneously, where M is equal to the number of states (Montvilas, 1995). The corresponding states at time moments $M + N = 5 + 20 = 25$ are presented in Table 2.

In Fig. 1 the mapping results are presented, where at the first $M = 5$ time moments the state vectors mapped simultaneously are denoted by mark $\times$ with an index that means the time moment number, and the state vectors mapped sequentially are denoted by mark + with the respective index.

The results of mapping show that the power station has five working states. The marks gathered on the screen into five separate places. During this experiment two damages at 11 and 24 time moments (frequency of 51Hz and 49Hz respectively) and two undesirable states at 15 and 17 time moments (voltage of 230V and 210V respectively) were simulated. As we see on Fig. 1, the undesirable states are indicated on the screen: the marks 15 and 17 are out of the place representing the 3-ed stable state, however the more important cases, the damages, were not revealed. This situation could be explained easy. The

Table 1

The values of parameters P at $1 \div 25$ time moments

TIME	PARAMETERS					
MOMENTS	P_1	P_2	P_3	P_4	P_5	P_6
1	50.00	220.0	307.0	508.0	604.0	206.0
2	50.01	220.2	302.0	501.0	603.0	205.0
3	50.00	220.8	304.0	502.0	607.0	208.0
4	50.08	221.0	305.0	505.0	609.0	204.0
5	50.03	220.6	308.0	507.0	608.0	202.0
6	50.04	220.4	307.0	507.7	604.0	206.0
7	50.01	220.2	307.2	508.0	604.0	206.3
8	50.02	220.0	307.0	508.2	604.0	206.1
9	50.06	220.0	307.1	507.9	603.9	205.8
10	50.08	220.2	302.0	501.0	603.0	205.1
11	**51.00**	220.0	302.1	500.9	603.1	204.9
12	50.08	220.0	302.2	501.2	603.2	205.0
13	50.06	220.6	302.2	501.5	603.0	205.5
14	50.04	221.0	304.2	502.0	607.0	208.0
15	50.02	**230.0**	304.2	502.3	607.3	208.1
16	50.00	220.0	304.0	502.0	607.0	208.2
17	50.00	**210.0**	303.8	501.8	606.4	207.9
18	50.01	218.0	304.5	504.7	608.7	203.9
19	50.00	219.0	305.5	504.9	609.1	204.1
20	49.99	220.0	304.9	505.0	609.2	204.0
21	49.98	220.2	304.8	505.2	609.0	204.0
22	49.92	220.2	307.9	507.0	609.2	202.0
23	49.92	220.0	308.0	507.5	608.0	202.5
24	**49.00**	220.0	309.0	507.5	607.2	202.0
25	50.00	220.0	308.1	507.0	608.0	201.9

Table 2

The states of PS at the time moments $M + N = 5 + 20 = 25$

MAPPING	SIMULTANEOUS					SEQUENTIAL					
MARK	×					+					
STATE	1	2	3	4	5	1	2	3	4	5	Damaged
TIME MOMENTS	1	2	3	4	5	6,7, 8, 9	10,12, 13	14,15, 16, 17	18,19, 20, 21	22,23, 25	11,24

changing of voltage from 220V to 230V or 210V means the changing of correspondent parameter value at ± 10, while the changing of frequency from 50Hz to 51Hz or 49Hz means the changing of corresponding parameter value only ± 1. Such a small changing is allowed for each parameter describing the state: on the Fig.1 the marks are gathered not

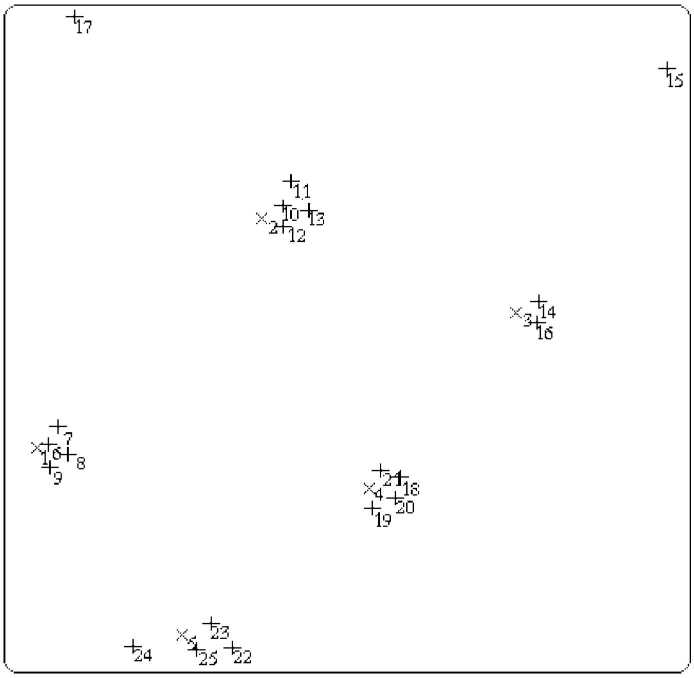

Fig. 1. The view on the PC screen of the ends of mapped vectors of PS for 25 time moments.

in one point, but according states they are distributed in some local places characterising the corresponding states of power station. The distances in L-dimensional and (after mapping) two-dimensional space are geometrically proportional to parameters numerical value and their changes, and the changing of ± 1 is involved into usual changing of parameters. Therefore the mapping did not reveal the damages of the power station. The other problem is that the parameters describing the states can be of any physical nature, so corresponding calculate parameters preserve *various* dimensions which anyway are lost after the mapping.

3. Solution of the Problems

To solve mentioned above problems it is necessary to make the changes of parameters to be more sensitive for representing in the space independently on their numerical values, and, in addition, dimensionless. The sequential nonlinear mapping has a peculiarity: it begins to work having only few vectors and later works sequentially by receiving each new vector of parameters. So at the beginning we can not know neither an average of parameter nor its dispersion. We can know or assign only the range of it. Hence this preprocessing of data could be done in such a way. Let the parameter P changes in range from $P_{\max}$ till $P_{\min}$. Independently of absolute value of $P_{\max}$ and $P_{\min}$ it is possible to make the range of *calculation* parameter S from $S_{\max}$ till $S_{\min}$ where, for simplicity, taking $S_{\max} = 1$ and $S_{\min} = 0$ (Spath, 1980). Then we can write

$$\frac{S - S_{\min}}{P - P_{\min}} = \frac{S_{\max} - S_{\min}}{P_{\max} - P_{\min}}, \tag{1}$$

and

$$S = \frac{(P - P_{\min})(S_{\max} - S_{\min})}{P_{\max} - P_{\min}} + S_{\min}. \tag{2}$$

Now, taking into account that we assumed to be $S_{\max} = 1$ and $S_{\min} = 0$, finally we can write

$$S = \frac{P - P_{\min}}{P_{\max} - P_{\min}}. \tag{3}$$

The calculation parameter S became a nondimensional one.

Thus, for each parameter we have to know its current value P and its range: $P_{\max}, P_{\min}$. For our example they are shown in the Table 3.

Now it is possible to calculate the parameters $S_1 \div S_6$ using formula (3). The calculation result is shown in Table 4.

If the value of some parameter P randomly run from its range then the value of corresponding calculation parameter S will be more than 1 or less than 0 (negative), but it is no matter, the mapping will be made as usual. The result of mapping of calculation parameter's vectors is presented in Fig. 2.

In Fig. 2 it is possible to see the real situation. At the time moments 11 and 24 the power station was in unusual (damage) states.

Choosing $P_{\max}$ and $P_{\min}$ we can make the mapping more or less sensitive for each parameter P. Let's make the mapping more sensitive for changing of parameter P_1. This case we have to use the smaller range for it. Let the range of parameter P_1 be $P_{1\max} - P_{1\min} = 50.2 - 49.8 = 0.4$. Then the Table 4 will change narrowly: at the time moment 11 the value of $S_1 = 3$ and at the time moment 24 $S_1 = -2$ respectively. The values S_1 at other time moments will change only a little. In Fig. 3 we see the situation when the mapping became more sensitive for changing of parameter P_1: the marks at the time moments 11 and 24 run away from the places representing the second and the fifth states respectively in major distances than that of Fig. 2. And finally, this pre-processing of data can be useful and for *simultaneous* nonlinear mapping.

Table 3

The ranges of each parameters

Parameters	P_1	P_2	P_3	P_4	P_5	P_6
$P_{\max}$	50.5	230.0	309.0	508.2	609.2	208.2
$P_{\min}$	49.5	210.0	302.0	500.9	603.0	201.9
$P_{\max} - P_{\min}$	1.0	20.0	7.0	7.3	6.2	6.3

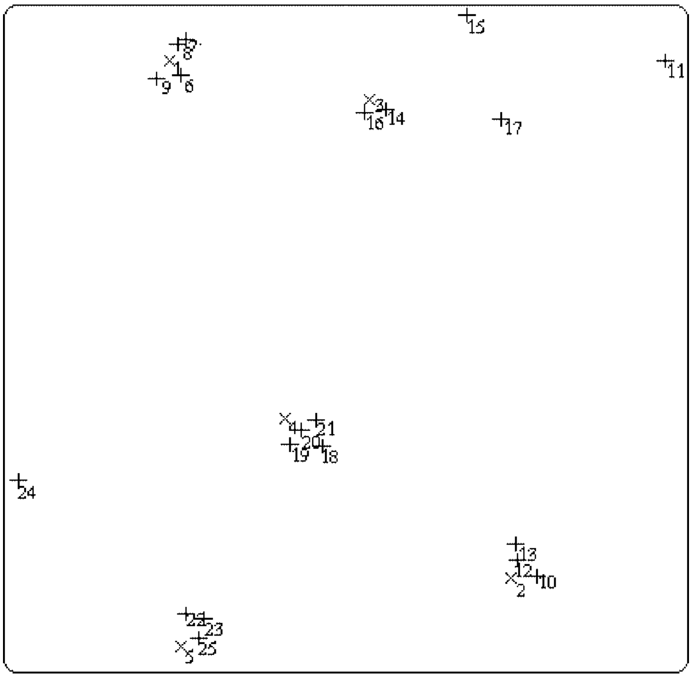

Fig. 2. The view of mapped calculation parameter's vectors.

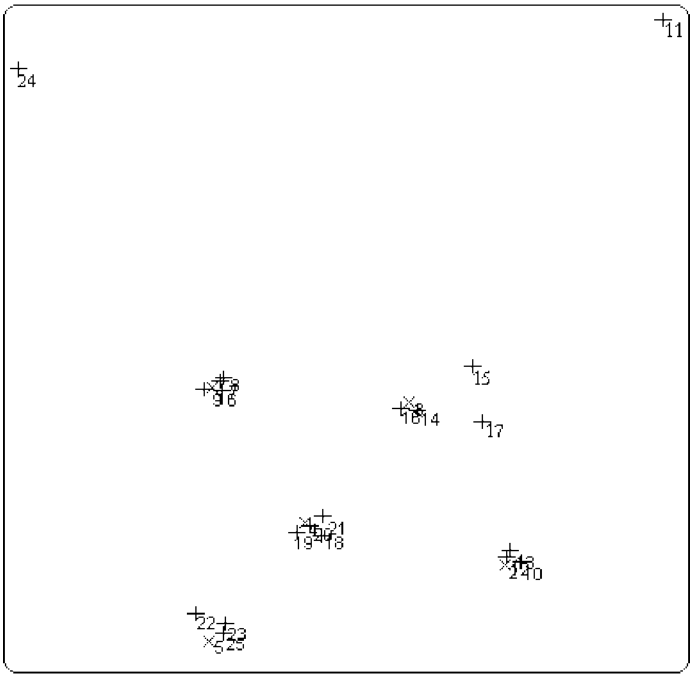

Fig. 3. The view of mapped vectors when the mapping is more sensitive for P_1 than that of Fig. 2.

Table 4

The values of calculation parameters S at $1 \div 25$ time moments

TIME MOMENTS	CALCULATION PARAMETERS S_1	S_2	S_3	S_4	S_5	S_6
1	0.50	0.50	0.714	0.973	0.161	0.651
2	0.51	0.51	0.000	0.014	0.000	0.492
3	0.50	0.54	0.289	0.150	0.645	0.968
4	0.58	0.55	0.428	0.562	0.968	0.333
5	0.53	0.53	0.857	0.836	0.806	0.016
6	0.54	0.52	0.714	0.931	0.161	0.651
7	0.51	0.51.	0.743	0.973	0.161	0.698
8	0.52	0.50	0.714	1.000	0.161	0.667
9	0.56	0.50	0.728	0.959	0.145	0.619
10	0.58	0.51	0.000	0.014	0.000	0.508
11	**1.50**	0.50	0.014	0.000	0.016	0.476
12	0.58	0.50	0.028	0.041	0.032	0.492
13	0.56	0.53	0.028	0.082	0.000	0.571
14	0.54	0.55	0.314	0.150	0.645	0.968
15	0.52	**1.00**	0.314	0.192	0.693	0.984
16	0.50	0.50	0.285	0.150	0.645	1.000
17	0.50	**0.00**	0.257	0.123	0.548	0.952
18	0.51	0.40	0.357	0.520	0.919	0.317
19	0.50	0.45	0.500	0.548	0.984	0.349
20	0.49	0.50	0.414	0.562	1.000	0.333
21	0.48	0.51	0.400	0.589	0.968	0.333
22	0.42	0.51	0.843	0.836	1.000	0.016
23	0.42	0.50	0.857	0.904	0.806	0.0965
24	**−0.50**	0.50	1.000	0.904	0.677	0.048
25	0.50	0.50	0.871	0.836	0.806	0.000

4. Discussions

The sequential nonlinear mapping can be successfully used for sequential detection the dynamic systems states. It enables to identify the current state, its changes and to detect the undesirable states or even damages of the dynamic systems. However there are the situations when parameter describing the state has to be constant, and a relatively small its change causes a damage of the dynamic system so it has to be watched. In order to make the method sensitive for the small changing of parameter it is necessary to pre-process the data using formula (2) or (3). After that the method becomes sensitive for every changing of parameters and it can be used for supervising of any dynamic system because the method reveals all undesirable states of dynamic systems and their damages. It is possible to change the sensitivity of the method by changing the range of corresponding parameters.

References

Lee R.C.T., J.R. Slangle, H. Blum (1977). A triangulation method for the sequential mapping of points from N-space to *two*-space. *IEEE Trans. on Computers*, Vol. **c-26**(3), 288–292.

Montvilas A.M. (1995). On sequential nonlinear mapping for data structure analysis. *Informatica*, **6**(2), 225–232.

Montvilas A.M. (1999a). Processing of information for supervision and control of technological processes. In *Proceedings of the IFAC Workshop*, Vienna, 2–4 Dec. 1999, PERGAMON, pp. 39–43.

Montvilas A.M. (1999b). Issues for design of information system for supervision and control of dynamic systems. *Informatica*, **10**(3), 289–296.

Sammon J.W. (1969). A nonlinear mapping for data structure analysis. *IEEE Trans. on Computers,* Vol.**c-18**(5), 401–409.

Spath. H. (1980). *Cluster Analysis for Data Reduction and Classification of Objects*. Ellis Horwood, Chichester.

A.M. Montvilas is a doctor of technical sciences, a senior researcher of the Recognition Processes Department at the Institute of Mathematics and Informatics and an Associate Professor of the Radioelectronics Department at the Vilnius Gediminas Technical University. His research interests include processing of random processes, classification, detection of dynamic system states and their changes, discrete signal processing, sequential nonlinear mapping.

Duomenų netiesiniam atvaizdavimui išankstinis apdorojimas

Algirdas Mykolas MONTVILAS

Straipsnyje parodoma, kad prieš identifikuojant dinaminių sistemų būsenas, naudojantis duomenų vektorių, aprašančių sistemos būsenas, netiesiniu atvaizdavimu, reikalingas išankstinis specialus duomenų apdorojimas, tuo būdu išryškinant nors ir nežymius, bet labai svarbius dinaminių sistemų būsenas aprašančių parametrų pokyčius, kuriems sistema yra jautri. Jautrumą kuriam nors parametrui galima keisti keičiant atitinkamo parametro užduodamą diapazoną. Po išankstinio duomenų apdorojimo galima stebėti kompiuterio ekrane sistemos būsenas, nustatyti nepageidaujamas būsenas ir netgi sėkmingai konstatuoti sistemos gedimus. Pateikti pavyzdžiai.