ISSN 1392 - 1215   ELEKTRONIKA IR ELEKTROTECHNIKA. 2005. Nr. 1(57)

T 170 ELEKTRONIKA

Optimal Initial Conditions for Nonlinear Mapping of Multidimensional Signals

A. M. Montvilas

Institute of Mathematics and Informatics, A. Goštauto str. 12, LT–01108 Vilnius, Lithuania; e-mail: montvila@ktl.mii.lt
Vilnius Gediminas Technical University, Naugarduko str. 41, LT-03227 Vilnius, Lithuania

Introduction

Now signals usable for various types of communications are described by many parameters. Consequently these signals are presented in multidimensional space. The Sammon's method of simultaneous nonlinear mapping for data structure analysis is used for clustering of the signals [1]. The essence of the method is to preserve the inner structure of distances among the signals in multidimensional space after mapping them into two-dimensional space. The formula (1) (mapping error) is used as criteria of mapping quality because it reveals the largest product of the error and partial error [2]:

$$E = \frac{1}{\sum_{i<j}^{N} d_{ij}^{*}} \sum_{i<j}^{N} \frac{\left(d_{ij}^{*} - d_{ij}\right)^2}{d_{ij}^{*}}, \qquad (1)$$

where N is a number of L-dimensional vectors being mapped, d_{ij}^{*} - distance between i and j vectors in L-space, d_{ij} - distance in a lower-dimensional space (two-space).

For the first, signals on the plane, as initial conditions, are distributed randomly or along some more or less shifted diagonal. Formula (2) is used for correction of co-ordinates of the signals on plane during each iteration r.

$$y_{pq}(r+1) = y_{pq}(r) - F * \Delta_{pq}(r), \qquad (2)$$

where p=1,…,N; q=1,2; F is "*magic factor*" (0.3-0.4);

$$\Delta_{pq}(r) = \frac{\partial E(r)}{\partial y_{pq}(r)} \Bigg/ \left| \frac{\partial^2 E(r)}{\partial y_{pq}^2(r)} \right|. \qquad (3)$$

The mapping error (1) depends on a sort of initial conditions, and there was no way to choose optimal initial conditions because any nonlinear mapping algorithm often finds the local maximum of a functional that characterizes the mapping quality which is not global [3].

In the paper operation of mapping algorithm has been revealed and the optimal initial conditions have been proposed.

Visualization of Mapping Process

An experiment has been executed in order to show how points representing multidimensional signals in L-dimensional space are moving on the plane during iteration procedure. The experiment was executed with a small amount of vectors for imaginary.

Table 1. 15 vectors of the parameters of signals

Vect. No	Parameters 1	2	3	4	5	6
1	50.0	220.0	307.0	508.0	604.0	206.0
2	50.1	220.2	307.2	508.0	604.0	206.3
3	50.2	220.5	307.1	507.9	603.9	205.8
4	51.0	222.0	302.0	501.0	603.0	205.0
5	50.9	221.8	302.1	500.9	603.1	204.9
6	50.8	222.6	302.2	501.5	603.0	205.5
7	53.0	228.0	304.0	502.0	607.0	208.0
8	52.9	227.8	304.2	502.3	607.3	208.1
9	53.1	227.9	303.8	501.8	606.4	207.9
10	55.0	226.0	305.0	505.0	609.0	204.0
11	55.0	226.5	304.5	504.7	608.7	203.9
12	55.1	225.9	305.5	504.9	609.1	204.1
13	56.0	224.0	308.0	507.0	608.0	202.0
14	56.2	224.2	307.9	507.0	609.2	202.0
15	55.9	224.4	308.4	507.5	608.0	202.5

They are distributed to five classes (Table 2).

Table 2. 15 vectors of data distributed to five classes

Class	Vectors
1	1,2,3
2	4,5,6
3	7,8,9
4	10,11,12
5	13,14,15

For the first the data were mapped onto the plane using initial conditions chosen by a random way and r=500 iterations. The result of mapping is presented in Fig. 1. The mapping error was E=0.04714809.

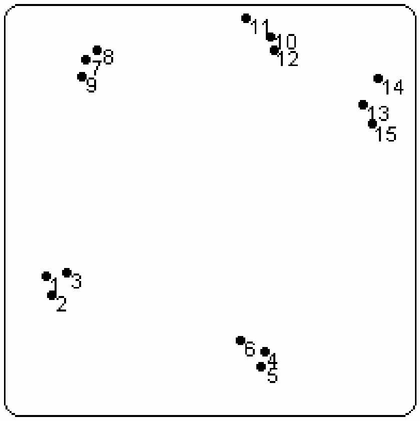

Fig. 1. Mapping of the data (Table 1)

After that the initial conditions were constructed in such a way that all points representing signals on the plane were distributed almost according their "right" positions (Fig. 1) except the 3-rd point which belongs to the first group of signals (including signals 2 and 3). It was put on the plane near the 14-th point (far from its group). Then mapping process was executed again but at small number of iterations in order to watch how point number 3 approaches to its group (Fig. 2). The mapping error was E=0.209099.

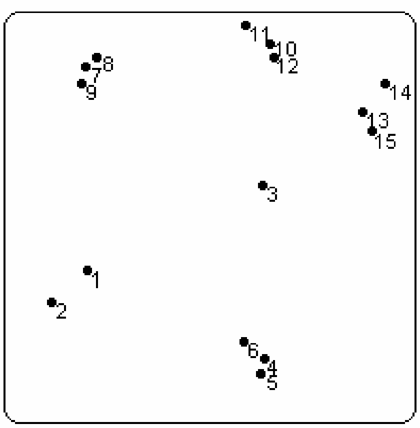

Fig. 2. Mapping of the data at small number of iterations

The results of mapping the data at more iterations are shown in Fig. 3. The mapping error was E=0.098837.

The 3-rd point approaches to its group but does it not very smart. It could be explain that another points had ideal initial conditions. They were almost in their "right" places on the plane at the very beginning of mapping. Consequently their distances structure on the plane corresponds to the distances structure in L-space and total mapping error is not big. Therefore the 3-rd point gets very small Δ_{pq}-correction of co-ordinates, because its calculation includes error of distances of another points, which are of small values already.

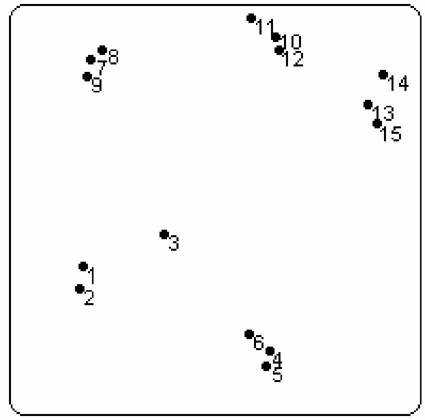

Fig. 3. Mapping of the data at more iterations

Consequently the points, which got "bad" initial conditions, are left behind their "right" places on the plane. This explains the mapping error variety using different initial conditions.

Another experiment with the same data and similar initial conditions has been executed. This case the 12-th point, which belongs to the fourth class of data, was put on the plane near the first point. For the first mapping process was executed at small number of iterations, as well. The result of mapping is presented in Fig. 4. The mapping error was E=0.344215.

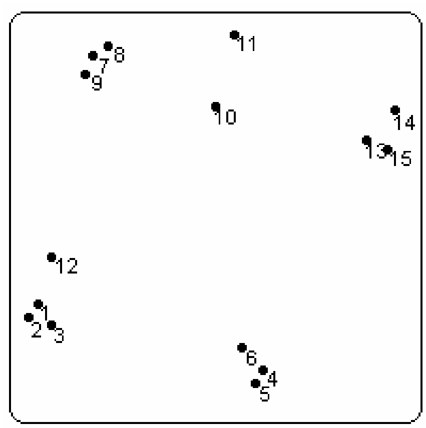

Fig. 4. View of mapped signals at small number of iterations

The view of mapped signals at more iterations is shown in Fig. 5. The mapping error was E=0.307327.

Eventually, the view of mapping at big number of iterations is shown in Fig. 6. The mapping error was E=0.060578. This case the point, which had "bad" initial conditions, deformed the total view of mapping (compare Fig. 6 and Fig. 1).

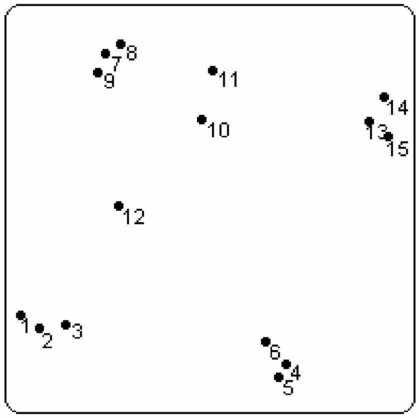

Fig. 5. View of mapped signals at more iterations

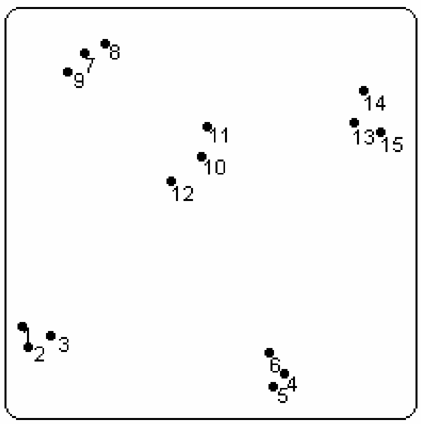

Fig. 6. View of mapped signals at big number of iterations

Optimal Initial Conditions

Initial conditions, as it was shown, is a “narrow” part of the nonlinear mapping process. It is possible to execute a mapping for ,suppose, a million various kinds of random initial conditions and as a mapping result to choose the case with the smallest mapping error E. But anyway not every point would have “good” initial conditions. There is no possibility to perform or check it.

Optimal initial conditions for simultaneous nonlinear mapping of the signals can be obtained using the *sequential* nonlinear mapping [4]. The sequential nonlinear mapping requires at the very beginning to map M initial vectors *simultaneously*, using Sammon’s algorithm. After that each sequentially receiving vector has to be mapped with respect to the first M vectors. Mapping error function E_j has to be minimized for each receiving vector $j=M+1,\ldots, M+N$, using formula

$$E_j = \frac{1}{\sum_{i=1}^{M} d_{ij}^X} \sum_{i=1}^{M} \frac{\left(d_{ij}^X - d_{ij}^Y\right)^2}{d_{ij}^X}, \quad j = M+1,\ldots, M+N; \quad (4)$$

where d_{ij}^X - distance between i and j vectors in the L-space, d_{ij}^Y - distance on the plane.

In [5] this sequential method and Sammon’s simultaneous one were compared according ability to map the data onto the plane, mapping accuracy and a mapping time. It was showed that sequential nonlinear mapping has slightly bigger total mapping error but needs incomparably less calculation time. In [6] the sequential nonlinear mapping was completely investigated.

In the issue regardless of that the sequential nonlinear mapping often has slightly bigger mapping error than simultaneous one but the sequential mapping gives the *optimal* initial conditions for simultaneous one because every point on the plane after sequential mapping is lying very close to its “right” place and there is no one point with “bad” initial conditions. In order to confirm that, numerous experiments have been executed with wide variety of sorts of signals, which differ according number of groups and number of vectors. Some of the results are presented in Table 3, where mapping errors of simultaneous mapping with random initial conditions (Simultan 1), sequential mapping (Sequential) and simultaneous mapping with optimal initial conditions (Simultan 2) are given.

Table 3. Mapping errors for 20 various cases

Exp No	Mapping errors		
	Simultan 1	Sequential	Simultan 2
1	0.049527	0.046792	0.044167
2	0.004849	0.005424	0.004294
3	0.004426	0.005424	0.004294
4	0.023269	0.026552	0.023185
5	0.026389	0.026551	0.023185
6	0.025812	0.678499	0.023218
7	0.049663	0.005424	0.004294
8	0.049527	0.436872	0.004375
9	0.049753	0.307663	0.004549
10	0.026389	0.027561	0.024472
11	0.025781	0.027549	0.024578
12	0.026524	0.264530	0.024317
13	0.004635	0.005424	0.004294
14	0.004416	0.005424	0.004294
15	0.004587	0.007279	0.004387
16	0.004729	0.005424	0.004294
17	0.004777	0.005413	0.004291
18	0.004514	0.762025	0.004305
19	0.004696	0.005424	0.004294
20	0.004500	0.005419	0.004285

Analyzing data of Table 2 we see that mapping errors of “Simultan 2” are always less than that of “Simultan 1”.

There was the case (7-th row) where sequential mapping error was less than “Simultan 1” error. It means that this case “Simultan 1” had especially “bad” initial conditions for many points.

Conclusions

There is no way to choose optimal initial conditions for the Sammon's nonlinear mapping of multidimensional signals onto the plane by distributing points on the plane randomly. The points with "bad" initial conditions do not reach their "right" places on the plane after mapping and this situation increases the total mapping error. The *sequential* nonlinear mapping, which gives a slightly bigger mapping error but needs incomparable less calculation time, is the only way to get the *optimal* initial conditions for the Sammon's nonlinear mapping.

References

1. **Sammon J. W.** A Nonlinear Mapping for Data Structure Analysis // IEEE Trans. on Computers.-1969.-Vol. c-18(5).– P. 401–09.
2. **Duda R. O., Hart P. E.** Pattern Classification and Scene Analysis. – New York, London, Sydney, Toronto: John Wiley & Sons, 1973.
3. **Dzemyda G.** Clustering of Parameters on the Basis of Correlations: a Comparative Review of Deterministic Approaches // Informatica. – 1997. – 8(1). – P. 83–118.
4. **Montvilas A. M.** On Sequential Nonlinear Mapping for Data Structure Analysis // Informatica. – 1995. – 6(2). – P. 225–232.
5. **Montvilas A. M.** Sequential Nonlinear Mapping versus Simultaneous One // Informatica. – 2002. – 13(3). – P. 333–343.
6. **Montvilas A. M.** Investigation of Sequential Mapping of Multidimensional Data // Elektronika ir elektrotechnika. – Kaunas: Technologija, 2003. – Nr.6(48). – P.7– 2.

Pateikta spaudai 2004 05 20

A. M. Montvilas. Optimalios pradinės daugiamačių signalų netiesinio atvaizdavimo sąlygos // Elektronika ir elektrotechnika. – Kaunas: Technologija, 2005. – Nr. 1(57). – P. 24–27.

Ryšiuose naudojami signalai yra aprašomi daugeliu parametrų, taigi jie vaizduojami daugiamatėje erdvėje. Signalams atvaizduoti taškais plokštumoje, o vėliau jiems klasterizuoti taikomas Sammono vienalaikio netiesinio atvaizdavimo plokštumoje metodas. To metodo esmė – išsaugoti vidinę atstumų tarp signalų parametrų vektorių struktūrą po jų atvaizdavimo. Atvaizdavimo kokybę nusako atvaizdavimo klaida, kuri labai priklauso nuo pradinių sąlygų. Pradinės sąlygos parenkamos atsitiktinai išdėsčius taškus plokštumoje arba pagal įstrižainę, arba kaip nors kitaip, tačiau nėra būdų *optimalioms* pradinėms sąlygoms parinkti, todėl atvaizdavimo kokybę nusakantis funkcionalas dažnai randa lokalinį minimumą, kuris nėra globalus. Darbe ištirtas atvaizdavimo procesas ir pasiūlytas *nuoseklaus* atvaizdavimo metodas optimalioms pradinėms sąlygoms nustatyti. Nuosekliojo atvaizdavimo klaida kiek didesnė negu vienalaikio, bet, jį pritaikius, taškai jau pačioje pradžioje optimaliai išdėstomi plokštumoje, o jų padėtis vėliau patikslinama Sammono metodu, duodančiu mažesnę atvaizdavimo klaidą. Gausūs eksperimentai tai patvirtina. Il. 6, bibl. 6 (anglų kalba; santraukos lietuvių, anglų ir rusų k.).

A. M. Montvilas. Optimal Initial Conditions for Nonlinear Mapping of Multidimensional Signals // Electronics and Electrical Engineering. – Kaunas: Technologija, 2005. – No. 1(57). – P. 24–27.

Complicated signals used in telecommunication are described by many parameters, hence they are presented in multidimensional space. Sammon's method of simultaneous nonlinear mapping onto the plane is used for clustering of the signals. The essence of the method is to preserve the inner structure of distances among the vectors of parameters of signals in multidimensional space after mapping them onto the plane. The quality of mapping is characterized by mapping error which very depends on initial conditions. The initial conditions are chosen on the plane randomly or along diagonal. There is no way to choose *optimal* initial conditions, therefore mapping algorithm often finds the local maximum of a functional that characterizes the mapping quality which is not global. In the paper operation of mapping has been revealed and *sequential* mapping has been proposed for choosing of optimal initial conditions. The sequential mapping gives slightly bigger mapping error then that of simultaneous one but it gives optimal initial conditions for simultaneous method, which maps the signal later with less mapping error. Numerous experiments confirm that. Ill. 6, bibl. 6 (in English; summarries in Lithuanian, English and Russian).

А. М. Монтвилас. Оптимальные начальные условия для нелинейного отображения многомерных сигналов // Электроника и электротехника. – Каунас: Технология, 2005. – № 1(57). – С. 24–27.

Сложные сигналы, используемые в связи, характеризуются многими параметрами и, таким образом, представляются в многомерном пространстве. Для отображения сигналов точками на плоскости и последующей их кластеризации применяется метод Саммона одновременного нелинейного отображения на плоскости. Сущность метода заключается в том, чтобы сохранить внутреннюю структуру расстояний между векторами параметров сигналов после их отображения на плоскости. Качество отображения характеризует ошибка отображения, которая сильно зависит от начальных условий. Начальные условия задаются случайным образом распределив точки на плоскости или по диагонали. Однако не существует способ для задания *оптимальных* начальных условий и функционал, определяющий качество отображения, часто находит локальный минимум, который не является глобальным. В работе исследован сам процесс отображения и предложен *последовательный* метод отображения для установки оптимальных начальных условий. Последовательное отображение дает несколько большую ошибку, чем одновременное отображение, но в результате оно оптимально распределяет точки на плоскости, положение которых впоследствии уточняет метод Саммона, дающий меньшую ошибку отображения. Многочисленные эксперименты это подтверждают. Ил. 6, библ. 6 (на английском языке; рефераты на литовском, английском и русском яз.).

DOI: 10.5755/j02.eie.10328