ISSN: 1803-3814

STRESS MEASURES IN SOM LEARNING

Zuzana Krbcova[1*], Jaromir Kukal[1]

[1]University of Chemistry and Technology in Prague
Faculty of Chemical Engineering
Department of Computing and Control Engineering
Technicka 5, 166 26 Prague 6
Czech Republic

Abstract: *Various stress measures can be used in generalized version of Sammon's mapping. Kohonen SOM with iterative or batch learning is a standard tool for data self-organization, too. Our method applies stress functions to pattern relationships in SOM and converts batch learning to discrete optimization task. Due to NP–completeness of SOM learning, optimization heuristics have to be used. Simulated annealing making use of Lévy flights is the recommended heuristics for this task.*

Keywords: *SOM, metric space, stress function, optimization heuristics*

1 Introduction

Objects represented by matrix of mutual distances are mapped into a structure of self organizing map so as to minimize stress functions. Optimization heuristic is applied to the finding minimum of stress measures. The following notation is used in this paper: let $\mathbb{A}$, $\mathbb{B}$ be non-empty finite sets representing input and output data spaces. In many applications $\mathbb{A} = \mathbb{R}^M$, $\mathbb{B} = \mathbb{R}^m$, where $M > m$ and Euclidean metrics is used. However our aim is to design a novel method using SOM [1, 2] topology. The output structure of SOM [3, 4] can be represented by graph $\mathcal{G}$, where $\mathcal{G} = \langle \mathbb{V}, \mathbb{E} \rangle$ is undirected connected graph consisting of vertex set $\mathbb{V} = \mathrm{V}(\mathcal{G})$, edge set $\mathbb{E} = \mathrm{E}(\mathcal{G}) \subset \binom{\mathbb{V}}{2}$, and cardinality of $\mathrm{V}(\mathcal{G})$ is H. The distance between two vertices $k, l \in \mathbb{V}$ is measured by natural metric $d_\mathrm{G}(k, l)$, which is defined as the length of shortest path connecting vertices k and l. Therefore, in our case $\mathbb{B} = \mathrm{V}(\mathcal{G})$. Let $\langle \mathbb{A}, \delta \rangle$, $\langle \mathrm{V}(\mathcal{G}), d_\mathrm{G} \rangle$ be two metric spaces. Let C be the number of patterns in $\mathbb{A}$. Square matrix $\mathbf{D} = (D_{ij})_{i,j=1}^{C}$ represents the distances between each pair of elements $\mathbf{a}_i, \mathbf{a}_j \in \mathbb{A}$, thus $D_{ij} = \delta(\mathbf{a}_i, \mathbf{a}_j)$. Similarly, square matrix $\mathbf{L} = (L_{ij})_{i,j=1}^{C}$ represents the distances between the corresponding mapped elements of set $\mathrm{V}(\mathcal{G})$ as $L_{ij} = d_\mathrm{G}(\mathbf{b}_{p_i}, \mathbf{b}_{p_j})$, where vector $\mathbf{p} \in \{1, \cdots, H\}^C$ is a vector assigning i-th element of $\mathbb{A}$ to p_i-th element of $\mathbb{B}$ and will be the subject of optimization.

2 Portfolio of Stress Functions

Metric multidimensional scaling (MMDS) includes a stress function which can be generalized by function $\mathrm{E}_{\mathrm{MMDS}}$. The form of function $\mathrm{E}_{\mathrm{MMDS}}$ is

$$E_{\mathrm{MMDS}} = \frac{1}{S} \sum_{i<j} (L_{ij} - D_{ij})^2 w_{ij},$$

where $S > 0$ is normalization scalar and $w_{ij} \geq 0$ is weight of D_{ij} in general. We use particular stress functions of MMDS [5]:

- Raw-stress: $E_{\mathrm{RS}} = \sum\limits_{i<j} (L_{ij} - D_{ij})^2$
- Sammon: $E_{\mathrm{SA}} = \frac{1}{\sum\limits_{i<j} D_{ij}} \sum\limits_{i<j} \frac{(L_{ij} - D_{ij})^2}{D_{ij}}$

and several stress functions of Bregmanised MMDS [6, 7]:

- Itakura-Saito: $E_{\mathrm{IS}} = \sum\limits_{i<j} \left(\frac{L_{ij}}{D_{ij}} - \ln \frac{L_{ij}}{D_{ij}} - 1 \right)$
- Reciprocal: $E_{\mathrm{RE}} = \sum\limits_{i<j} \left(\frac{1}{L_{ij}} - \frac{2}{D_{ij}} + \frac{L_{ij}}{D_{ij}^2} \right)$
- Inverse quadratic: $E_{\mathrm{IQ}} = \sum\limits_{i<j} \left(\frac{1}{L_{ij}^2} - \frac{3}{D_{ij}^2} + \frac{2L_{ij}}{D_{ij}^3} \right)$
- Extended Sammon: $E_{\mathrm{ESA}} = \sum\limits_{i<j} \left(L_{ij} \ln \frac{L_{ij}}{D_{ij}} - L_{ij} + D_{ij} \right)$.

3 Self Organizing Mapping and Its Stress

An input metric space is frequently linear with Euclidean distance. Comparison of the Euclidean distance D_{ij} with natural graph distance $L_{ij} = d_G(\mathbf{b}_{p_i}, \mathbf{b}_{p_j})$ enforced graph distance scaling by formula $L_{ij}^{\text{new}} = \alpha L_{ij}$, where $\alpha > 0$ is defined as minimizer of each stress function using explicit formulas

- E_{RS}: $\alpha = \frac{\sum\limits_{i<j} L_{ij} D_{ij}}{\sum\limits_{i<j} L_{ij}^2}$
- E_{RE}: $\alpha = \sqrt{\frac{\sum\limits_{i<j} L_i}{\sum\limits_{i<j} \frac{L_{ij}}{D_{ij}}}}$
- E_{SA}: $\alpha = \frac{\sum\limits_{i<j} L_{ij}}{\sum\limits_{i<j} \frac{L_{ij}^2}{D_{ij}}}$
- E_{IQ}: $\alpha = \left(\frac{\sum\limits_{i<j} L_{ij}^{-2}}{\sum\limits_{i<j} \frac{L_{ij}}{D_{ij}^3}} \right)^{1/3}$
- E_{IS}: $\alpha = \frac{(C-1)C}{2 \sum\limits_{i<j} \frac{L_{ij}}{D_{ij}}}$
- E_{ESA}: $\alpha = \exp\left(-\frac{\sum\limits_{i<j} L_{ij} \ln \frac{L_{ij}}{D_{ij}}}{\sum\limits_{i<j} L_{ij}} \right)$.

This data scaling does not change distance proportions. The resulting stress functions are then optimized with respect to real parameter α. Therefore mixed integer optimization tasks are converted to discrete optimization ones with unknown partition $\mathbf{p}$.

4 Partition Optimization

Heuristic optimization technique with embedded Lévy flights [8] was used for integer minimization of stress and correlation measures. The technique is similar to Fast Simulated Annealing (FSA) [9] but Cauchy distribution is used in mutation operator instead of an acceptance decision. We also used fixed temperature according to [8].

Supposing optimization domain is $\mathbb{D} = \{1, \cdots, H\}^C$, we minimize the objective function $\mathrm{f} : \mathbb{D} \to \mathbb{R}_0^+$. The optimum value is denoted as $\mathrm{f}^* = \min\limits_{\mathbf{p} \in \mathbb{D}} \mathrm{f}(\mathbf{p})$, and, therefore, we find any point satisfying

$$\mathbf{p}_{\text{opt}} \in \arg\min_{\mathbf{p} \in \mathbb{D}} \mathrm{f}(\mathbf{p}).$$

Integer heuristics description uses three basic terms: uniform distribution on any point set is denoted as $\mathrm{U}(\cdots)$, unit sphere around origin in $\mathbb{R}^C$ is denoted as $\mathbb{S}_{C-1}$ and $\mathrm{perturb}(\mathbf{p}, \mathbb{D})$ is a function which may push point $\mathbf{p}$ into $\mathbb{D}$ by using boundary reflection. The algorithm of random descent with Lévy flight [9] mutation has three parameters: dimensionless temperature $T > 0$, (sub)optimum value $\mathrm{f}^+ \geq \mathrm{f}^*$ and N as the maximum number of $\mathrm{f}(\mathbf{p})$ evaluations.

The algorithm begins from random point $\mathbf{p}_0 \sim \mathrm{U}(\mathbb{D})$, $k = 0$ and continues for $\mathrm{f}(\mathbf{p}_k) > \mathrm{f}^+ \wedge k < N$ in loop:

- $d_k = \tan \frac{\pi}{2} \eta_k$, $\eta_k \sim \mathrm{U}([0, 1])$ as Lévy flight length
- $\mathbf{v}_k \sim \mathrm{U}(\mathbb{S}_{C-1})$ as Lévy flight direction
- $\mathbf{p}_{\text{trial}} = \left[\mathbf{p}_k + T d_k \mathbf{v}_k\right]$ as free point in $\mathbb{Z}^C$
- $\mathbf{p}_{\text{new}} = \mathrm{perturb}(\mathbf{p}_{\text{trial}}, \mathbb{D})$ as new point in $\mathbb{D}$
- $\mathbf{p}_{k+1} = \begin{cases} \mathbf{p}_{\text{new}} & \text{for} \quad \mathrm{f}(\mathbf{p}_{\text{new}}) < \mathrm{f}(\mathbf{p}_k) \\ \mathbf{p}_k & \text{for} \quad \mathrm{f}(\mathbf{p}_{\text{new}}) \geq \mathrm{f}(\mathbf{p}_k) \end{cases}$ as stringent decision rule
- $k = k + 1$.

This heuristic is relatively simple, but due to Lévy flights [10] it has useful performances.

5 Referential Measures

Pearson ρ_{P}, Spearman ρ_{S} and Kendall τ correlation coefficients are used as referential SOM quality measures. They are defined as follows

$$\rho_{\text{P}} = \frac{\sum\limits_{i<j} \left(D_{ij} - \overline{D}\right)\left(L_{ij} - \overline{L}\right)}{\sqrt{\sum\limits_{i<j} \left(D_{ij} - \overline{D}\right)^2} \sqrt{\sum\limits_{i<j} \left(L_{ij} - \overline{L}\right)^2}},$$

$$\rho_S = 1 - \frac{6 \sum\limits_{i<j} r_{ij}^2}{n(n^2-1)},$$

$$\tau = \frac{n_c - n_d}{\sqrt{(n_0 - n_1)(n_0 - n_2)}},$$

where $\overline{D} = \frac{1}{n} \sum\limits_{i<j} D_{ij}; \overline{L} = \frac{1}{n} \sum\limits_{i<j} L_{ij}$; $n = C(C-1)/2$; $r_{ij} = x_{ij} - y_{ij}$, i.e. x_{ij} represents rank of sorted Euclidean distances D_{ij}, i.e. y_{ij} is rank of sorted graph distances L_{ij}; n_c is the number of concordant pairs; n_d is the number of discordant pairs; $n_0 = n(n-1)/2$; n_1 is the number of ties including the vector of all distances D_{ij}; and n_2 is the number of ties including the vector of all distances L_{ij}. Finally, quantities $E_\mathrm{P} = 1 - \rho_\mathrm{P}, E_\mathrm{S} = 1 - \rho_\mathrm{S}$ and $\mathrm{E_K} = 1 - \tau$ are subjects of discrete minimization.

6 Case Study: SOM with Hexagonal Topology

The first aim of our computer experiments was to investigate reliability and time complexity of SOM learning for $\mathbb{A} = \mathbb{B} = \mathrm{V}(\mathcal{G})$. The optimum value of any stress or referential function is $\mathrm{f}^* = 0$, which is easy to prove. We use hexagonal SOM topology of nineteen vertices ($H = 19$) for seven data points ($C = 7$). Various testing data sets are shown in Fig. 1. Adequate optimization domain $\mathbb{D}$ consist of $19^7 = 893{,}871{,}739$ states which are difficult to evaluate systematically. After application of simulated annealing making use of Lévy flights for $T = 19, \mathrm{f}^+ = 0, N = 30000$, we obtain reliabilities and time complexities for various measures and pattern sets as demonstrated in Tables 1 and 2. The three most reliable measures are $E_\mathrm{P}\,(19.2\%), E_\mathrm{K}\,(19.0\%)$ and $E_\mathrm{RS}\,(17.4\%)$. Time complexity of SOM learning is small for E_RS (10526), E_P (10848) and E_ESA (10988). Using Feoktistov criterion [11] $FEO = MNE/REL$, where MNE is the average number of function evaluation for successful trials and REL is the percentage of successful to total trials, the best measure is E_P with $FEO = 56500$ (Table 3).

Table 1: Reliability [%] of SOM learning for $\mathbb{A} = \mathrm{V}(G)$

Measure	Number of axes					Average
	0	1	2	6 S	6 F	
E_RS	12	12	40	14	9	17.4
E_SA	11	10	21	13	7	12.4
E_IS	9	8	17	13	8	11.0
E_RE	9	8	12	14	8	10.2
E_IQ	5	7	8	7	5	6.4
E_ESA	11	12	28	16	8	15.0
E_P	14	15	39	15	13	19.2
E_S	14	8	41	10	6	15.8
E_K	14	9	44	15	13	19.0

Table 2: Time complexity of SOM learning for $\mathbb{A} = \mathrm{V}(G)$

Measure	Number of axes					Average
	0	1	2	6 S	6 F	
E_RS	9925	10307	7497	16326	8577	10526
E_SA	15286	11408	12534	17749	13513	14098
E_IS	10684	10771	16883	13860	11141	12668
E_RE	14527	16751	14323	12459	14226	14457
E_IQ	12007	16813	15092	15749	19160	15752
E_ESA	8545	14206	8339	15182	8670	10988
E_P	9997	12515	8640	13925	9162	10848
E_S	11049	15445	9263	14026	6822	11321
E_K	11207	15457	9533	9745	10845	11357

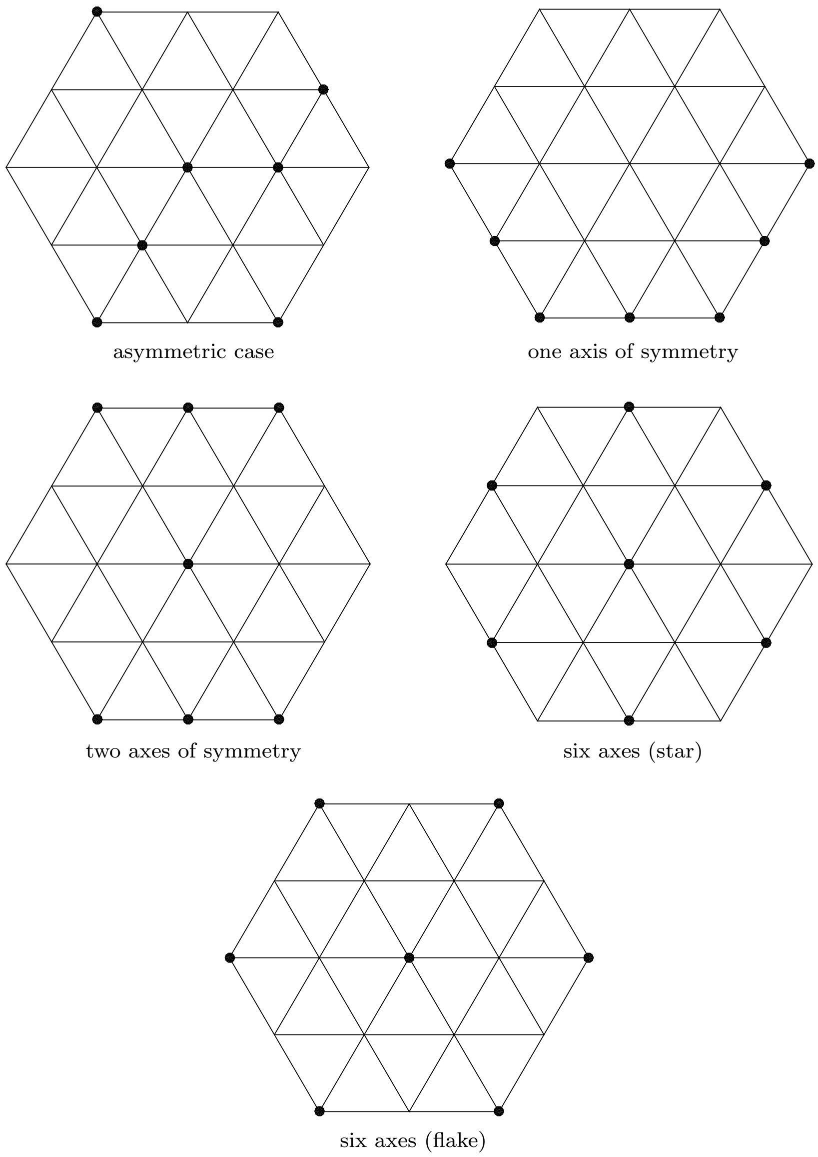

Figure 1: Graphical representation of input sets

Table 3: Feoktistov criterion

Measure	Mapping from	
	$\mathbb{R}^2$	$\mathrm{V}(G)$
E_{RS}	60494	68402
E_{SA}	113694	89718
E_{IS}	115164	123698
E_{RE}	141735	174800
E_{IQ}	246125	173803
E_{ESA}	73253	98342
E_{P}	56500	67791
E_{S}	71652	97085
E_{K}	59774	101955

However, a typical application of SOM is mapping from $\mathbb{A} = \mathbb{R}^M$ into $\mathbb{B} = \mathrm{V}(\mathcal{G})$ where in our case $M = 2$. The optimum stress value is non-negative in general. Data patterns were the same as in the previous case, but their distances were calculated via Euclidean distance from coordinates. The optimum values of measures are included in Table 4. In the second case, reliability and time complexity of SOM learning was studied for $\mathbb{A} = \mathbb{R}^2$. After application of simulated annealing making use of Lévy flights for $T = 19$, $H = 19, \mathrm{f}^+ = \mathrm{f}^*, N = 30000$ and $C = 7$, we obtain reliabilities and time complexities for various measures and pattern sets as collected in Tables 5 and 6. The three most reliable measures are E_{RS} (16.4 %), E_{P} (15.8 %) and E_{SA} (14.2 %). Time complexity of SOM learning is small for E_{P} (10711), E_{RS} (11218) and E_{ESA} (11801). The best measure is again E_{P} with $FEO = 67791$.

Table 4: Optimum values of SOM measures for $\mathbb{A} = \mathbb{R}^2$

Measure	Number of axes				
	0	1	2	6 S	6 F
E_{RS}	0.5394	0.6026	0.7566	1.1028	0.8616
E_{SA}	0.0037	0.0039	0.0040	0.0046	0.0049
E_{IS}	0.0450	0.0445	0.0430	0.0434	0.0452
E_{RE}	0.4430	0.0389	0.0388	0.0247	0.0285
E_{IQ}	0.0662	0.0500	0.0402	0.0209	0.0260
E_{ESA}	0.1008	0.1076	0.1204	0.1529	0.1394
E_{P}	0.0124	0.0093	0.0100	0.0279	0.0201
E_{S}	0.0233	0.0146	0.0551	0.0000	0.0339
E_{K}	0.0552	0.0505	0.1205	0.0000	0.0742

Table 5: Reliability [%] of SOM learning for $\mathbb{A} = \mathbb{R}^2$

Measure	Number of axes					Average
	0	1	2	6 S	6 F	
E_{RS}	11	11	39	11	10	16.4
E_{SA}	10	14	25	12	10	14.2
E_{IS}	11	8	14	9	6	9.6
E_{RE}	10	5	10	10	5	8.0
E_{IQ}	6	7	13	6	6	7.6
E_{ESA}	12	9	20	13	6	12.0
E_{P}	13	11	26	16	13	15.8
E_{S}	11	8	20	11	15	13.0
E_{K}	13	8	16	15	15	13.4

Table 6: Time complexity of SOM learning for $\mathbb{A} = \mathbb{R}^2$

Measure	Number of axes 0	1	2	6 S	6 F	Average
E_{RS}	9357	11005	5636	12093	17999	11218
E_{SA}	5701	13073	15878	15024	14024	12740
E_{IS}	15207	10075	13316	7742	13036	11875
E_{RE}	12591	15159	15848	12569	13754	13984
E_{IQ}	14178	16417	16208	7747	11497	13209
E_{ESA}	14260	12554	12553	14024	5615	11801
E_{P}	12029	9837	9012	9931	12747	10711
E_{S}	12772	13749	9560	13547	13479	12621
E_{K}	16352	11780	13562	9782	16836	13662

7 Conclusion

Six stress measures and three correlation measures were used for batch SOM learning via FSA making use of Lévy flights. Experimental results for mappings from 2D (planar graph and $\mathbb{R}^2$) to hexagonal grid can be generalized as follows:

In the case of perfect mapping from hexagonal into the same topology recognized E_{P} (19.2 %) most reliable for heuristic optimization, while E_{RS} (10526) measure has less possible time complexity. From the multicriteria decision making theory point of view, both E_{P} and E_{RS} are just two Pareto optimal choices. Using Feoktistov criterion, the best value $FEO = 56500$ was obtain for Pearson correlation E_{P}. Except for E_{IQ}, the reliability of measures higher than 10 % and their time complexities are comparable.

Similar results were obtained for mapping from $\mathbb{R}^2$ into hexagonal topology with just two Pareto optimal choices, namely E_{RS} with the best reliability (16.4 %) and E_{P} with the smallest possible time complexity (10711). Pearson correlation E_{P} is again the best compromise approach with $FEO = 67791$. Except for E_{IS}, E_{RE}, E_{IQ}, all the measures have higher reliability than 10 % with comparable time complexity.

Therefore, we recommend E_{RS}, E_{SA}, E_{ESA} stress measures and E_{P}, E_{S}, E_{K} correlations for batch SOM learning via FSA making use of Lévy flights.

Acknowledgement: This paper has been supported financially from specific university research MSMT No 21-SVV/2018 and grant GA17-05840S "Multicriteria Optimization of Shift-Variant Imaging System Models" of the Czech Science Foundation.

References

[1] Kohonen, T.: Self-Organizing Maps, third edn. Springer-Verlag, Berlin, Heidelberg, New York (2001)

[2] Abe, T., Kanaya, S., Kinouchi, M., Kudo, Y., Mori, H., Matsuda, H., Carlos, D.C., and Ikemura, T.: Gene classification method based on batch-learning SOM. Genome Informatics **10**, 314–315 (1999)

[3] Kanaya, S., Kinouchi, M., Abe, T., Kudo, Y., Yamada, Y., Nishi, T.,Mori, H., Ikemura, T.: Analysis of codon usage diversity for bacterial genes with a self-organizing map (SOM): characterization of horizontally transferred genes with emphasis on the E. coli O157 genome. Gene **276**, 89–99 (2001)

[4] Mojzeš M., Kukal J., Tran, V. Q., Jablonský, J.: Performance comparison of heuristic algorithms via multi-criteria decision analysis. In: R. Matousek (ed.) Proceedings of 17th International Conference on Soft Computing – MENDEL 2011, pp. 244–251. Brno University of Technology, VUT Press, Brno (2011)

[5] Sammon, J.W.: A nonlinear mapping for data structure analysis. IEEE Transaction on Computers **C-18**(5), 401–409 (1969)

[6] Sun, J., Crowe, M., Fyfe, C.: Extending metric multidimensional scaling with Bregman divergences. Pattern Recognization **44**(5), 1137–1154 (2011)

[7] Sun, J., Fyfe, C., Crowe, M.: Extending Sammon mapping with Bregman divergences. Information Sciences **187**(1), 72–92 (2012)

[8] Kukal, J.: SOM in Metric Space. Neural Network World **14**(6), 469–488 (2004)

[9] Szu, H.: Fast simulated annealing. In: Denker J.S. (ed.) Proceeding of the AIP Conference on Neural Networks for Computing, pp. 420-425. American Institute of Physics, New York (1986)

[10] Viswanathan, G.M., Raposo, E.P., Da Luz, M.G.E.: Lévy flights and superdiffusion in the context of biological encounters and random searches. Physics of Life Reviews **5**(3), 133–150 (2008)

[11] Feoktistov, V.: Differential evolution, *In Search of Solutions*. Springer, Boston, MA (2006)