10.24425/acs.2019.131233

Archives of Control Sciences
Volume 29(LXV), 2019
No. 4, *pages 699–717*

Anomalous and traditional diffusion modelling in SOM learning

RADEK HREBIK and JAROMÍR KUKAL

The traditional self organizing map (SOM) is learned by Kohonen learning. The main disadvantage of this approach is in epoch based learning when the radius and rate of learning are decreasing functions of epoch index. The aim of study is to demonstrate advantages of diffusive learning in single epoch learning and other cases for both traditional and anomalous diffusion models. We also discuss the differences between traditional and anomalous learning in models and in quality of obtained SOM. The anomalous diffusion model leads to less accurate SOM which is in accordance to biological assumptions of normal diffusive processes in living nervous system. But the traditional Kohonen learning has been overperformed by novel diffusive learning approaches.

Key words: self organization, Kohonen map, diffusion learning, anomalous diffusion, SOM

1. Introduction

The self organizing map (SOM) is a traditional tool for data analysis which transforms the data patterns from the input space into vertices of an undirected SOM graph with a given topology and unit length edges. The input patterns are from metric vector space in many applications. The parameters of SOM are the weights which are placed into vertices and are subject of learning. The Kohonen learning [19] is the first approach which is frequently used in many applications [4, 24, 27, 38]. The main disadvantage of this approach is in epoch based learning with necessity of learning parameter changes. This learning algorithm has also a weak biological motivation but there are many alternative approaches with better properties.

The authors are with FNSPE, CTU in Prague, Trojanova 13, 120 00, Prague 2, Czech Republic. The corresponding author is E. Hrebik, E-mail: Radek.Hrebik@fjfi.cvut.cz.

The authors would like to acknowledge the support of the research grant SGS 17/196/OHK4/3T/14. The second author also acknowledges the support of the OP VVV MEYS funded project CZ.02.1.01/0.0/0.0/16_019/0000765 Research Center for Informatics.

Received 6.12.2018.

The alternatives are strongly connected with brain physiology study about slow signal propagation in central nervous system. Nitric oxide (NO) is well known neurotransmitter in mammal brain due to its ability to diffuse isotropically in aqueous and lipid environments [32]. Using NO as an intercellular signaling molecule in the nervous system has been confirmed by many studies [12, 17]. The way of information transmission by neurons in both vertebrates and invertebrates have also been discussed by many authors [8, 11, 13, 29]. The intervening cellular or membrane structures are discussed in [23, 39]. The whole surface of the neuron is therefore potential release site for NO, in marked contrast to conventional transmitter release, which is restricted to the synaptic zone [16, 30, 36].

A lot of physiological studies served as a background model for the realisation of artificial self organisation systems. Lopez et. al [25, 26] developed two pure informatics models yielding from the simplification of NO dynamic. Models are not focused to exact physical description of diffusion process. Moreover, the spatial effect is modelled as multi–compartment discrete system in these studies.

Our previous study [18] has been focused on modelling of traditional diffusion [5, 6] and yields from primal neurophysiological studies [12, 17] to obtain adequate learning algorithm as simplification of real nitric oxide diffusion process. We also formulate alternative model based on the anomalous diffusion having another properties during learning process. Our aim is to compare Kohonen learning, diffusive learning with traditional diffusion and learning strategy based on anomalous diffusion in the case of single and multiple epoch learning of SOM.

The second section is oriented to two pudding models of NO diffusion. They are based on tradition or anomalous diffusion in infinite continuum where the neurons are placed as generators of concentration pulses. The various learning approaches of SOM are included in the third section which is focused on Kohonen learning, normal and anomalous diffusive learning. A list of traditional measures of SOM quality is included in the fourth section together with their time complexity analysis. Three case studies are presented in the fifth section which illustrates learning quality of proposed diffusive techniques on artificial data sample and traditional testing datasets – iris flower and wine quality tasks including conclusions in the last section.

2. Free diffusion in $\mathbb{R}^d$

Supposing the diffusion of nitric oxide is the main slow learning phenomenon of neural systems we have to model this process in physical, chemical and mathematical sense. We decided to analyse only such models of diffusion with chemical reaction which offer analytical solution in infinite continuum of given dimension. There are only two cases which satisfy previous condition: traditional diffusion for exponent $\alpha = 2$ and anomalous diffusion constrained to case when $\alpha = 1$.

The *pudding model* description begins with remembering of basic facts. Let $m, n, H \in \mathbb{N}$ be number of patterns, pattern dimensionality and number of SOM neurons [14]. The individual patterns are $\mathbf{x}_j \in \mathbb{R}^n$, where $j = 1, \ldots, m$ and form the pattern set $\mathcal{S} = \{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$. The fixed positions of individual neurons in continuum are $\mathbf{p}_i \in \mathbb{R}^N$ for $i = 1, \ldots, H$ and reflects the topology of SOM [28] which is subject of network design. The diffusion process in continuum can be easily expressed using matrix $\mathbf{D} \in (\mathbb{R}_0^+)^{H \times H}$ of distances $d_{i,j} = \|\mathbf{p}_i - \mathbf{p}_j\|_2$. These mutual distances inderictly express the topology of SOM. In pudding SOM the neuron distances are not coinstraned to integeres what enables better space mapping. Therefore, the resulting SOM is invariant to transition and rotation of its structure. Let $\Delta t > 0$ be learning period and the diffusion in continuum will be studied only in discrete time $t_k = k \cdot \Delta t$, where $k \in \mathbb{N}_0$. The result of SOM learning is the system of weights [3] $\mathbf{w}_i \in \mathbb{R}^n$, where $, i = 1, \ldots, H$ of course. We begin with random weights setting $\mathbf{w}_i(0)$. The weights evolve during learning process and their values in time t_q are denoted as $\mathbf{w}_i(q)$, where $q \in \mathbb{N}_0$. The *pudding model* is based on substrate concentrations in neurons and given time. Being prepared to SOM learning we have to study the concentration profile first using single and complete activation procedures.

The case of traditional diffusion ($\alpha = 2$) has been discussed in [18] and the main results introduce us into the formalism of diffusion process.

2.1. Traditional approach

The slow information transfer in the nervous system can be modelled as diffusion process [6] with first order chemical reaction [9]. The reactant is generated by single neuron activity [21] and the diffusion process [40] spread the substance in the neuron neighbourhood. Our model is based on the second Fick's law [2] of diffusion which is modified by kinetics of pseudo-monomolecular [35] chemical reaction. The neuron activity can be modelled as Dirac impulse in given time. The main advantage of these simplifications is in the existence of analytical solution which can be obtained as follows.

Let $N \in \mathbb{N}$ be space dimension, $\mathbf{y} \in \mathbb{R}^N$ be point coordinate, $D, t, \lambda > 0$ be the diffusion coefficient, time and the rate constant of a chemical reaction. The free diffusion of reacting substrate of concentration $\mathrm{c} : \mathbb{R}^N \to \mathbb{R}_0^+$ is driven by partial differential equation

$$\frac{\partial \mathrm{c}(\mathbf{y}, t)}{\partial t} = D \nabla^2 \mathrm{c}(\mathbf{y}, t) - \lambda \mathrm{c}(\mathbf{y}, t) \tag{1}$$

with initial condition

$$\mathrm{c}(\mathbf{y}, 0_+) = \delta(\mathbf{y}), \tag{2}$$

where $\delta : \mathbb{R}^N \to \mathbb{R}_0^+$ is Dirac function.

The fundamental solution of (1) is

$$c(\mathbf{y}, t) = \frac{1}{(4\pi Dt)^{N/2}} \cdot \exp\left(-\frac{\|\mathbf{y}\|_2^2}{4Dt}\right) \cdot \exp(-\lambda t). \tag{3}$$

Due to system linearity, time and space invariance of (1) we can use the fundamental solution to study of multi-neuron system with sequential activities.

Being prepared to SOM learning we have to study the concentration profile first.

2.1.1. Single activation

The pudding SOM learning is based on the activation of a single neuron. We will study j-th neuron which is supposed to be active in time t_k. Therefore, formally $j = \varphi_k$. But it is not necessary to study the substrate concentration in any point. The learning is based only on the concentration (3) in neuron points. The concentration in time t_q is

$$c(\mathbf{y}, \mathbf{p}_j, t_q) = \frac{1}{(4\pi D(t_q - t_k))^{N/2}} \cdot \exp\left(-\frac{\|\mathbf{y} - \mathbf{p}_j\|_2^2}{4D(t_q - t_k)}\right) \cdot \exp\left(-\lambda(t_q - t_k)\right) \tag{4}$$

for $q > k$. The formula can be simplified to

$$c(\mathbf{p}_i, \mathbf{p}_j, t_q) = \frac{1}{(4\pi D(q-k)\Delta t)^{N/2}} \cdot \exp\left(-\frac{d_{i,j}^2}{4D(q-k)\Delta t}\right) \cdot \exp\left(-\lambda(q-k)\Delta t\right). \tag{5}$$

After the substitution $a = 4D\Delta t > 0$, $b = \lambda\Delta t > 0$ we obtain resulting activation formula

$$c(\mathbf{p}_i, \mathbf{p}_j, t_q) = (\pi a(q-k))^{-N/2} \cdot \exp\left(-\frac{d_{i,j}^2}{a(q-k)} - b(q-k)\right). \tag{6}$$

When $\min(d_{i,j} \geqslant 1)$, then we suggest to use $a = 1, b = 1/10$ for the first experiments as will be demonstrated in next sections.

2.1.2. Complete activation

The SOM learning is based on substrate concentrations in q-th step in the time t_q. This concentration is a result of previous activation sequence $\varphi_1, \varphi_2, \ldots, \varphi_{q-1}$ using single activation model (6). Due to linearity of (1) we can use the additivity principle and directly calculate the cumulative concentration in the i-th neuron and step q

$$c_{i,q} = \sum_{k=1}^{q-1} c(\mathbf{p}_i, \mathbf{p}_{\varphi_k}, t_q - t_k) = \frac{1}{(\pi a)^{N/2}} \cdot \sum_{k=1}^{q-1} \frac{\exp\left(-\frac{d_{i,\varphi_k}^2}{a(q-k)} - b(q-k)\right)}{(q-k)^{N/2}}. \tag{7}$$

Resulting formula consists of all concentration information which is necessary for the SOM learning. Therefore, the concentration $c_{i,q}$ is only a function of activation history, SOM topology and parameters a, b. The difference between single and complete activation is depicted on figure 1 for normal diffusive learning. The concentration of substrate is very high in the neighborhood of last winning neuron of the history. But the history is result of learning which will be studied in the next section.

2.2. Anomalous diffusion

As rarely observed in nature, the anomalous diffusion [33] is a more complex alternative to the traditional one. Both formulation and solution of models with anomalous diffusion are very complicated and not trivial except of case when $\alpha = 1$ which is sometimes called ballistic diffusion. We will formulate the model in general form first. Let $1 \leqslant \alpha < 2$, $D_\alpha > 0$ be anomalous exponent and diffusion coefficient. The free anomalous diffusion of reacting substrate of concentration $\mathrm{c} : \mathbb{R}^N \to \mathbb{R}_0^+$ is driven by partial differential equation

$$\frac{\partial \mathrm{c}(\mathbf{y}, t)}{\partial t} = D_\alpha \nabla^{(\alpha)} \mathrm{c}(\mathbf{y}, t) - \lambda \mathrm{c}(\mathbf{y}, t) \tag{8}$$

with initial condition

$$\mathrm{c}(\mathbf{y}, 0_+) = \delta(\mathbf{y}), \tag{9}$$

where $\delta : \mathbb{R}^N \to \mathbb{R}_0^+$ is Dirac function.

The explicit solution is obtainable only for $\alpha = 1$. The fundamental solution for $\lambda = 0$ is probability distribution function of multivarietal Cauchy distribution [22] for scale $\gamma = D_1 t$

$$\mathrm{c}(\mathbf{y}, t) = \frac{\Gamma\left(\dfrac{N+1}{2}\right)}{\Gamma(1/2)\pi^{N/2}} \cdot \frac{D_1 t}{\left(D_1^2 t^2 + \|\mathbf{y}\|_2^2\right)^{(N+1)/2}}. \tag{10}$$

Using shift theorem [7] of Laplace transform we obtain the general solution

$$\mathrm{c}(\mathbf{y}, t) = \frac{\Gamma\left(\dfrac{N+1}{2}\right)}{\Gamma(1/2)\pi^{N/2}} \cdot \frac{D_1 t \cdot \exp(-\lambda t)}{\left(D_1^2 t^2 + \|\mathbf{y}\|_2^2\right)^{(N+1)/2}} \tag{11}$$

and therefore,

$$\mathrm{c}(\mathbf{y}, \mathbf{p}, t_q) = \frac{\Gamma\left(\dfrac{N+1}{2}\right)}{\Gamma(1/2)\pi^{N/2}} \cdot \frac{D_1 (t_q - t_k) \cdot \exp(-\lambda (t_q - t_k))}{\left(D_1^2 (t_q - t_k)^2 + \|\mathbf{y} - \mathbf{p}\|_2^2\right)^{(N+1)/2}}, \tag{12}$$

$$c(\mathbf{p}_i, \mathbf{p}_j, t_q) = \frac{\Gamma\left(\dfrac{N+1}{2}\right)}{\Gamma(1/2)\pi^{N/2}} \cdot \frac{D_1(q-k)\Delta t \cdot \exp(-\lambda(q-k)\Delta t)}{\left(D_1^2(q-k)^2(\Delta t)^2 + d_{i,j}^2\right)^{(N+1)/2}}. \tag{13}$$

After the substitution $a = D_1\Delta t > 0$, $b = \lambda \cdot \Delta t > 0$ we obtain

$$c(\mathbf{p}_i, \mathbf{p}_j, t_q) = \frac{a\Gamma\left(\dfrac{N+1}{2}\right)}{\Gamma(1/2)\pi^{N/2}} \cdot \frac{(q-k)\cdot\exp(-b(q-k))}{\left(a^2(q-k)^2 + d_{i,j}^2\right)^{(N+1)/2}}, \tag{14}$$

$$c_{i,q} = \sum_{k=1}^{q-1} \frac{a\Gamma\left(\dfrac{N+1}{2}\right)}{\Gamma(1/2)\pi^{N/2}} \cdot \sum_{k=1}^{q-1} \frac{(q-k)\cdot\exp(-b(q-k))}{\left(a^2(q-k)^2 + d_{i,j}^2\right)^{(N+1)/2}}. \tag{15}$$

The learning efficiency depends on concentration profiles. Therefore, it is useful to compare substrate concentrations in the case of anomalous diffusion with the normal case. Results of single and complete activation are depicted in Fig. 2 for anomalous diffusive learning. As seen in Figs. 1 and 2 the substrate is more spread in the case of anomalous diffusion and the difference between minimal and maximal concentration are smaller. These differences between normal and anomalous diffusions influence the SOM learning algorithm and the main aim of this paper is to decide what kind of diffusion is more suitable for new type of diffusive SOM learning.

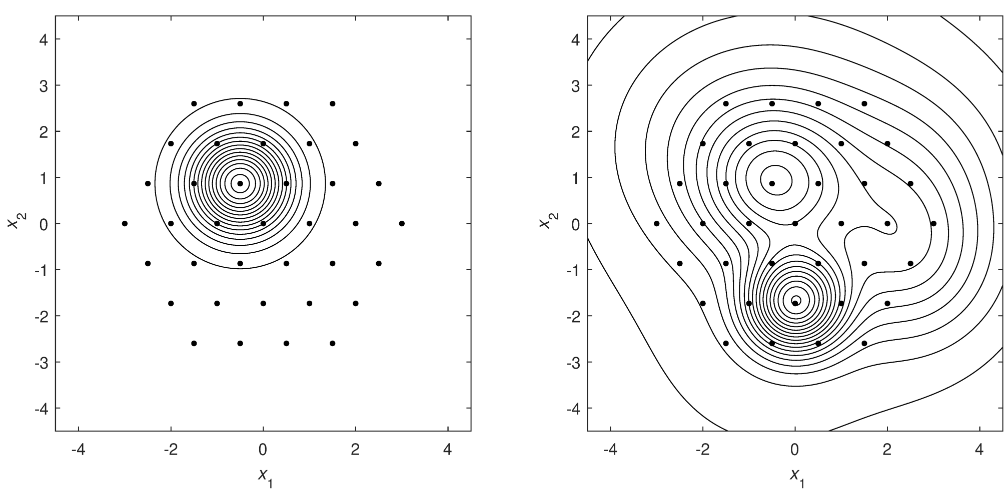

Figure 1: Concentration profile for normal diffusive learning after single (left) and complete (right) activation ($N = 2, a = 1, b = 1/10, H = 37, q = 100$)

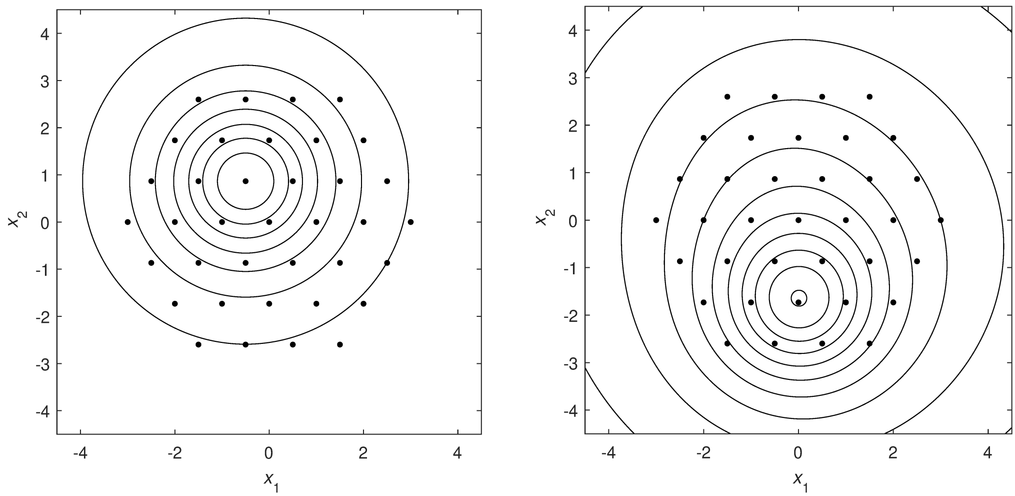

Figure 2: Concentration profile for anomalous diffusive learning after single (left) and complete (right) activation ($N = 2, a = 1, b = 1/10, H = 37, q = 100$)

3. SOM learning approach

There are many approaches how to perform modelling of self organisation. They can be directly inspired by anatomy and physiology of neuronal system or rather by other ideas which are easy to realize. Our research is inspired by the *pudding model* of atom in physics [21, 37], where the nucleus of atoms are supposed as points (raisins) in the electron continuum (pudding). In the case of self organisation we will place individual neurons instead of the atom nucleus into the continuum which would transfer the information in the system. But the main question is which model of nitric oxide diffusion is more suitable for SOM learning and whether the diffusive learning over-perform the traditional Kohonen in learning quality. In case of SOM we differ between three main approaches.

3.1. Kohonen learning

Kohonen network maps input vectors (patterns) of arbitrary dimension N onto a discrete map with 1 or 2 dimensions. One of main expected results is that patterns close to one another in the input space are to be close to one another in the map. This is called to be topologically ordered. Kohonen network is composed of a grid of output units and N input units. The input pattern is fed to each output unit. The input lines to each output unit are weighted. These weights are initialised to small random numbers.

In case of Kohonen learning [20] we use rules as follows. The weight of i-th neuron is changed in q-th step by rule

$$\mathbf{w}_i(q) = \mathbf{w}_i(q-1) + \alpha(q) \cdot c_{i,q} \cdot (\mathbf{x}_q - \mathbf{w}_i(q-1)) \tag{16}$$

for $i = 1, \ldots, H$, $\mathbf{x}_q \sim \mathrm{U}(\mathcal{S})$ is uniformly selected pattern from $\mathcal{S}$, $c_{i,q}$ is the substrate concentration according to (7) and $\alpha(q) > 0$ is the ageing function which is supposed to be non-increasing. The winner selection according to the Kohonen rule [20]

$$\varphi_q \in \arg\min_{k=1,\ldots,H} \|\mathbf{x}_q - \mathbf{w}_k\|_2. \tag{17}$$

As in the traditional SOM learning we have to initialize the weights [1] and use appropriate ageing strategy. We generate the initial weights from the multivariate Gaussian distribution as

$$\mathbf{w}_i(0) \sim \mathrm{N}(\mathrm{E}\mathbf{X}, \mathrm{var}\mathbf{X}/100) \tag{18}$$

for $i = 1, \ldots, H$. The ageing function $\alpha(q)$ can be constant in the first experiments, but satisfying $\alpha(q) \cdot c_{i,q} \leqslant 1$ to avoid learning instability.

3.2. Normal and anomalous diffusive learning

Novel learning algorithm is completely devoted to Kohonen learning rules [20] as follows. The weight of i-th neuron is changed in q-th step by rule

$$\mathbf{w}_i(q) = \mathbf{w}_i(q-1) + \alpha(q) \cdot c_{i,q} \cdot (\mathbf{x}_q - \mathbf{w}_i(q-1)) \tag{19}$$

for $i = 1, \ldots, H$, $\mathbf{x}_q \sim \mathrm{U}(\mathcal{S})$ is uniformly selected pattern from $\mathcal{S}$, $c_{i,q}$ is substrate concentration according to (7) or (15) respectively and $\alpha(q) > 0$ is ageing function which is supposed to be non-increasing. The winner is also selected according to Kohenen rule [20] as

$$\varphi_q \in \arg\min_{k=1,\ldots,H} \|\mathbf{x}_q - \mathbf{w}_k\|_2. \tag{20}$$

The main difference between the traditional SOM learning [1] and our approaches is in the application of diffusive equations (1) and (8) which generate the concentration profiles (7) and (14). The learning feedback is driven by the winner index φ_q from (17) which is used in the next step of concentration calculations (7) and (14).

As in the traditional SOM learning we have to initialize the weights [1] and use appropriate ageing strategy. We recommend to generate the initial weights from the multivarietal Gaussian distribution as

$$\mathbf{w}_i(0) \sim \mathrm{N}(\mathrm{E}\mathbf{X}, \mathrm{var}\mathbf{X}/100) \tag{21}$$

for $i = 1, \ldots, H$. The ageing function $\alpha(q)$ can be constant in the first experiments, but satisfying $\alpha(q) \cdot c_{i,q} \leqslant 1$ to avoid learning instability.

4. SOM quality measures

There are two main problems in SOM learning. First of them is called butterfly structure when the patterns are mapped in SOM graph with higher topographic error. The second problem is in a low accuracy of self organisation when the weights of SOM are far from the pattern set and the quantization error is higher. We will specify these measures first. The basic way of quality measurement design is based on measuring distances. The Euclidean distance of points $\mathbf{x}, \mathbf{y}$ in $\mathbb{R}^n$ is denoted $\mathrm{d}(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_2$.

Using the pattern $\mathbf{x}_j$ we can investigate the distances to weights $\mathbf{w}_k$ and define winner as

$$\mathrm{win}(j) \in \arg\min_{k=1,\ldots,H} \mathrm{d}(\mathbf{x}_j - \mathbf{w}_k) \tag{22}$$

but the function $\mathrm{win}(j)$ is of stochastic nature due to possible distance equities. In some cases we found the winner but one i.e. the second winner which is defined as

$$\mathrm{win2}(j) \in \arg\min_{k\in\mathcal{M}_j} \mathrm{d}(\mathbf{x}_j - \mathbf{w}_k), \tag{23}$$

where $\mathcal{M}_j = \{1, \ldots, H\} \setminus \{\mathrm{win}(j)\}$.

Using distances and winners we can design traditional measures of various nature.

4.1. Distance penalization

The Quantization Error (QE) is traditionally related to all forms of vector quantization and clustering algorithms [34]. Using linear penalisation we directly penalise the distances between patterns and corresponding winner weights as

$$QE_1 = \sum_{j=1}^{m} \mathrm{d}(\mathbf{x}_j, \mathbf{w}_{\mathrm{win}(j)}). \tag{24}$$

The quadratic penalisation

$$QE_2 = \sum_{j=1}^{m} \mathrm{d}^2(\mathbf{x}_j, \mathbf{w}_{\mathrm{win}(j)}) \tag{25}$$

is also frequently used but has higher sensitivity to outliers.

4.2. Topographic error

General topographic rule is: if two objects are close in reality they must be closed also in the map. Using this principle topographic error (TE) [15] is defined as

$$TE = 1 - \frac{1}{m} \sum_{j=1}^{m} g_{\mathrm{win}(j),\mathrm{win2}(j)}, \tag{26}$$

where $\mathbb{G} \in \{0,1\}^{H\times H}$ is SOM topology matrix with $g_{u,v} = \mathrm{I}(\|\mathbf{p}_v - \mathbf{p}_v\|_2 \leqslant 1)$. The main advantage of TE is in its robustness to outliers.

4.3. Correlation based measures

The correlations between mutual distances of patterns and mutual distances of winner weights can be directly used as quality measures.

Let i, j be pattern indexes. The mutual pattern distances can be defined as $d_{i,j} = \mathrm{d}(\mathbf{x}_i, \mathbf{x}_j)$. The mutual distances of corresponding weights are
$\delta_{i,j} = \mathrm{d}(\mathbf{w}_{win(i)}, \mathbf{w}_{win(j)})$.

Finally, we obtain $m(m-1)/2$ pairs of corresponding distances and directly calculate Pearson correlation coefficient r, Spearmann ρ or Kendall τ coefficient as quality measure. The correlation coefficients are frequently declared as p-values of independence hypothesis H_0 to be comparable with significant level 0.05.

4.4. Time complexity of measures

The evaluations of QE_1, QE_2 and TE are very fast with time complexity $\mathrm{O}(mnH)$. The evaluation of correlation measures is more complex. The Pearson r has time complexity $\mathrm{O}(mnH + m^2)$ due to simple statistics over $m(m-1)/2$ distance pairs. The Spearmann ρ is complicated with pair sorting and its time complexity is $\mathrm{O}(mnH + m^2 \log(m))$. The Kendall τ is not recommended for large pattern sets due to time complexity $\mathrm{O}(mnH + m^4)$.

4.5. Composed quality measures

In our research we prefer the QE_1 as main optimal criterion. Due to sensitivity to outliers we use QE_2 only as supplementary. Due to higher time sentiment we do not apply the correlation measures. The TE can be interpreted as probability of topology saving in random graph. Comparing TE as p-value with critical level of α we can test the hypothesis H_0 whether the resulting topology is random. Therefore, $TE \leqslant \alpha$ indicates significant topology of SOM mapping. Accepting only significant topology we can focus only to QE_1 and avoid butterfly effect.

5. Anomalous and traditional diffusion modelling in SOM learning

To compare three different SOM learning processes we use also three learning approaches based on different number of learning epochs. We evaluate the quality of SOMs for one, two and three epoch learning strategy in case of traditional Kohonen, diffusive and anomalous SOM.

In case of traditional Kohonen learning strategy we used following variable ranges:

- Number of epochs $E \in \{1, 2, 3\}$.
- Learning rate $\alpha \in \{5, 2, 1, 0.75, 0.5, 0.4, 0.3, 0.2, 0.1, 0.05, 0.01, 0.001\}$.
- Radius $R = \{10, 5, 3, 2, 1, 0.75, 0.5, 0.4, 0.3, 0.2, 0.1, 0.05\}$.
- Learning steps $N_k = 50000$.

In case of diffusive learning strategies we used following variable ranges:

- Number of epochs $E \in \{1, 2, 3\}$.
- Diffusive parameter $a = \{5, 4, 3, 2.5, 1, 0.75, 0.6, 0.5, 0.4, 0.3, 0.25, 0.2, 0.15, 0.1, 0.01, 0.001\}$.
- Kinetic parameter $b = \{0, 0.1, 0.2, 0.5, 1\}$.
- Learning rate $\alpha_0 = \{10, 5, 3, 2, 1, 0.75, 0.5, 0.3, 0.1, 0.05, 0.01, 0.001\}$.
- Learning steps $N_k = 50000$.

In all cases we performed ten experiments and the best solution satisfying $TE < 0.05$ and $QE_1 = \min$ has been found every setting.

Using this general methodology we test and compare results of three learning approaches on three different datasets. First dataset is our own artificial dataset to easily see the performance of tested approach. The second dataset is represented by traditional iris dataset and the third one is represented by wine quality dataset.

5.1. Case study I: Uniform data sample

As the first intuitive way of comparison we used artificial dataset. The dataset is useful for testing the SOM quality. We generated 10 000 randomly distributed points in the neighbourhood of 19 nodes of SOM in hexagonal topology. Individual patterns were generated as Gaussian mixture of 19 cases with mean value corresponding to node positions and standard deviation $\sigma = 0.2$.

5.1.1. Single epoch learning

First, we obtain the results for single epoch learning i.e. $E = 1$. The best results of single epoch learning are collected in Table 1 as quantitative errors (QE_1, QE_2) and topographic errors together with the best parameter setting. The best results were obtained for diffusion SOM followed by anomalous SOM and the worse

Table 1: Case study I: Single epoch learning

Method	Parameters	QE_1	QE_2	TE
Kohonen	$R = 0.50$ $\alpha_0 = 0.30$	0.07689	0.08443	0.00000
Diffusion SOM	$\alpha = 2,\ b = 0.00$ $a = 0.40$ $\alpha_0 = 0.35$	0.07180	0.07622	0.00000
Anomalous SOM	$\alpha = 1,\ b = 0.10$ $a = 0.30$ $\alpha_0 = 0.30$	0.07422	0.07894	0.03250

result has been obtained by Kohonen algorithm which is not recommended for single epoch learning.

The results of two epoch learning are included in Table 2 in the same meaning. In case of two epoch learning we still see the best results of Diffusion SOM but in case of Anomalous SOM and Kohonen we obtained comparable results.

Table 2: Case study I: Two epoch learning

Method	Parameters	QE_1	QE_2	TE
Kohonen	$R = (2.00, 0.50)$ $\alpha_0 = (0.5000, 0.1000)$	0.07214	0.07822	0.00000
Diffusion SOM	$\alpha = 2,\ b = 0.00$ $a = (2.00, 0.01)$ $\alpha_0 = (3.00, 0.001)$	0.06954	0.07442	0.00010
Anomalous SOM	$\alpha = 1,\ b = 0.00$ $a = (0.75, 0.10)$ $\alpha_0 = (0.60, 0.001)$	0.07278	0.07889	0.04210

Three epoch learning strategy confirms the power of Diffusion SOM as seen in Table 3. But Kohonen learning strategy brings better results than anomalous SOM.

This artificial example leads to following rules:

- All three approaches are able to reduce $TE < 0.05$.
- Optimal kinetic constant b in many optimal cases.
- The measures QE_1, QE_2 bring the same order of methods.

- Diffusion SOM overperformed anomalous SOM and Kohonen SOM learning in all cases.
- Kohonen method overperformed anomalous SOM in case of multi-epoch learning, i.e. $E \geqslant 3$.

Table 3: Case study I: Three epoch learning

Method	Parameters	QE_1	QE_2	TE
Kohonen	$R = (5.00, 3.00, 0.01)$ $\alpha_0 = (2.00, 0.50, 0.05)$	0.06977	0.07487	0.01500
Diffusion SOM	$\alpha = 2,\ b = 0.00$ $a = (0.20, 0.15, 0.10)$ $\alpha_0 = (3.00, 0.50, 0.001)$	0.06640	0.07050	0.00000
Anomalous SOM	$\alpha = 1,\ b = 0.00$ $a = (1.00, 0.50, 0.25)$ $\alpha_0 = (10.00, 1.00, 0.006)$	0.07150	0.07721	0.01990

Therefore, we will use two standardized datasets in next sections to demonstrate the real use of alternative methods. We compare the results for one and three epoch learning strategies.

5.2. Case study II: Iris flower task

We employ traditional iris flower classification task [10] to demonstrate the quality of SOM learning methods in the second case. The total number of 150 patterns of three classes (Iris setosa, Iris virginica, Iris versicolor) is described by four properties (sepal length, sepal width, petal length, petal width). We will compare only the results of 1 and 3 epoch learning strategies.

In the case of single epoch learning we collect the best results in Table 4 meanwhile the three epoch learning results are in Table 5.

The rules of optimal learning are same as in previous artificial example:

- All three approaches are able to reduce $TE < 0.05$.
- Optimal kinetic constant b in many optimal cases.
- Diffusion SOM overperformed anomalous SOM and Kohonen SOM learning in all cases.
- Anomalous diffusion brings worse results in case of QE_2 and single epoch learning than Kohonen SOM learning.
- Kohonen method overperformed anomalous SOM in case of multi-epoch learning, i.e. $E \geqslant 3$.

Table 4: Case study II: Single epoch learning

Method	Parameters	QE_1	QE_2	TE
Kohonen	$R = 0.75$ $\alpha_0 = 0.20$	0.22732	0.25794	0.03333
Diffusion SOM	$\alpha = 2,\ b = 1.00$ $a = 0.75$ $\alpha_0 = 0.75$	0.20627	0.23137	0.04000
Anomalous SOM	$\alpha = 1,\ b = 0.20$ $a = 0.50$ $\alpha_0 = 0.75$	0.21612	0.26052	0.04667

Table 5: Case study II: Three epoch learning

Method	Parameters	QE_1	QE_2	TE
Kohonen	$R = (1.00, 0.75, 0.50)$ $\alpha_0 = (0.30, 0.10, 0.01)$	0.18318	0.20630	0.02000
Diffusion SOM	$\alpha = 2, b = 0.00$ $a = (1.00, 0.75, 0.50)$ $\alpha_0 = (1.00, 0.50, 0.05)$	0.16611	0.18383	0.04000
Anomalous SOM	$\alpha = 1,\ b = 0.00$ $a = (1.00, 0.50, 0.25)$ $\alpha_0 = (2.00, 1.00, 0.01)$	0.21023	0.23457	0.04667

5.3. Case study III: Wine quality

Finally we use the traditional white wine quality task. Its dataset is represented by 4 898 patterns of 12 continuous attributes [31].

We compared only single and three epoch learning strategies again. In the case of single epoch learning we collect the best results in Table 6 meanwhile the three epoch learning results are in Table 7.

The rules of optimal learning are same as in previous artificial example:

- All three approaches are able to reduce $TE < 0.05$.
- Optimal kinetic constant b in many optimal cases.
- Diffusion SOM overperformed anomalous SOM and Kohonen SOM learning in QE_1.
- Kohonen learning achieves worse results in single epoch learning.

Table 6: Case study III: Single epoch learning

Method	Parameters	QE_1	QE_2	TE
Kohonen	$R = 5.00$ $\alpha_0 = 0.40$	19.58427	26.92369	0.03818
Diffusion SOM	$\alpha = 2,\ b = 0.50$ $a = 2.50$ $\alpha_0 = 0.75$	8.33177	12.93502	0.04512
Anomalous SOM	$\alpha = 1,\ b = 0.00$ $a = 2.50$ $\alpha_0 = 0.50$	12.29983	18.21313	0.04757

Table 7: Case study III: Three epoch learning

Method	Parameters	QE_1	QE_2	TE
Kohonen	$R = (5.00, 2.00, 1.00)$ $\alpha_0 = (5.00, 1.00, 0.01)$	8.64661	12.04015	0.04962
Diffusion SOM	$\alpha = 2,\ b = 0.20$ $a = (5.00, 3.00, 0.10)$ $\alpha_0 = (0.10, 0.50, 0.05)$	8.28362	12.73005	0.04226
Anomalous SOM	$\alpha = 1,\ b = 0.00$ $a = (5.00, 3.00, 1.00)$ $\alpha_0 = (0.10, 0.08, 0.05)$	8.47783	12.91590	0.04920

- Kohonen method overperformed anomalous SOM in QE_1 and QE_2 in case of three epoch learning.
- Kohonen method overperformed diffusion SOM in QE_2 only in case of three epoch learning.

6. Conclusions

Novel method of SOM learning based on anomalous diffusion has been developed and experimentally compared with SOM with normal diffusion and with Kohonen SOM learning. General aim was to decide whether the anomalous diffusion is able to improve the quality of SOM learning. The side effect of this study is in the optimal parameter setting which is easy to generalize. In all cases the SOM learning with normal diffusion overperformed the other approaches in case of QE_1, but the traditional Kohonen learning is worse than anomalous diffusion learning in QE_1 only in the case of single epoch learning. In the case of QE_2 criterion the diffusion SOM is the best choice only for the single epoch learning. The kinetic parameter b of both diffusion model can be set to zero in many cases which is strongly recommended choice. Moreover the behaviour in the case of artificial two–dimensional dataset is very similar to the behaviour on real datasets of higher dimensions. Our algorithm is nature inspired, slow diffusion of nitric oxide is its significant part and multi–epoch learning strategies are omitted as in the case of brain learning.

References

[1] E. Alonso: *Computational Neuroscience for Advancing Artificial Intelligence: Models, Methods and Applications: Models, Methods and Applications*. Premier reference source. Medical Information Science Reference, 2010.

[2] D. Brogioli and A. Vailati:. Diffusive mass transfer by nonequilibrium fluctuations: Fick's law revisited. *Phys. Rev. E*, **63**, 012105 (2000).

[3] W.L. Chang, L.M. Pang and K.M. Tay: Application of self-organizing map to failure modes and effects analysis methodology. *Neurocomputing*, **249** (2017), 314–320.

[4] J.A.F. Costa, A.P.V. Pinto, J.R. de Andrade and M.G. de Medeiros: Clustering of regional hdi data using self-organizing maps. In *2017 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*, (2017), 1–6.

[5] J. Crank: *The mathematics of diffusion*, J. Crank ed. Clarendon Press Oxford [England], 2nd ed. edition, 1975.

[6] E. Cussler: *Diffusion: Mass Transfer in Fluid Systems*. Cambridge Series in Chemical Engineering. Cambridge University Press, 2009.

[7] B. Davies: *Integral Transforms and Their Applications*. Texts in Applied Mathematics, Springer New York, 2002.

[8] G. Edelman and J. Gally: Nitric oxide: linking space and time in the brain. *Proceedings of the National Academy of Sciences*, **89**(24), (1992), 11651–11652.

[9] J. Espenson: *Chemical Kinetics and Reaction Mechanisms*. Advanced Chemistry Series, McGraw-Hill, 1995.

[10] R.A. Fisher: The use of multiple measurements in taxonomic problems. *Annals of eugenics*, **7**(2), (1936), 179–188.

[11] J.A. Gally, P.R. Montague, G.N. Reeke and G.M. Edelman: The no hypothesis: possible effects of a short-lived, rapidly diffusible signal in the development and function of the nervous system. *Proceedings of the National Academy of Sciences*, **87**(9), (1990), 3547–3551.

[12] J. Garthwaite, S.L. Charles and R. Chess-Williams: Endothelium-derived relaxing factor release on activation of nmda receptors suggests role as intercellular messenger in the brain. *Nature*, **336**(6197), (1988), 385–388.

[13] A. Gelperin: Nitric oxide mediates network oscillations of olfactory interneurons in a terrestrial mollusc. *Nature*, **369**(6475), (1994), 61–63.

[14] D. Graupe: *Deep Learning Neural Networks: Design and Case Studies*, 2016.

[15] L. Hamel: Som quality measures: An efficient statistical approach. In *Proceedings of the 11th International Workshop WSOM 2016*, pages 49–59, Houston. Springer, 2016.

[16] N.A. Hartell: Strong activation of parallel fibers produces localized calcium transients and a form of ltd that spreads to distant synapses. *Neuron*, **16**(3), (1996), 601–610.

[17] C. Hölscher: Nitric oxide, the enigmatic neuronal messenger: its role in synaptic plasticity. *Trends in neurosciences*, **20**(7), (1997), 298–303.

[18] R. Hrebik and J. Kukal: Diffusion modelling: Topographic error of som under control. *Soft Computing* (2018), page submitted.

[19] T. Kohonen: Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, **43**(1), (1982), 59–69.

[20] T. Kohonen: *Self-Organizing Maps*. Springer Series in Information Sciences. Springer Berlin Heidelberg, 2012.

[21] S. Kornblith, R.Q. Quiroga, C. Koch, I. Fried and F. Mormann: Persistent single-neuron activity during working memory in the human medial temporal lobe. *Current Biology*, **27**(7), (2017), 1026–1032.

[22] S. Kotz and S. Nadarajah: *Multivariate T-Distributions and Their Applications*. Cambridge University Press, 2004.

[23] J.R. Lancaster: Simulation of the diffusion and reaction of endogenously produced nitric oxide. *Proceedings of the National Academy of Sciences*, **91**(17), 1994), 8137–8141.

[24] A. Lavecchia: Machine-learning approaches in drug discovery: methods and applications. *Drug Discovery Today*, **20**(3), (2015), 318–331.

[25] P.F. Lopez, C.P.S. Araujo, P.G. Baez and G.S. Martin: Diffusion associative network: diffusive hybrid neuromodulation and volume learning. In *International Work-Conference on Artificial Neural Networks*, pp. 54–61, Springer, 2003.

[26] P.F. Lopez, P.G. Baez and C.P.S. Araujo: Nitric oxide diffusion and multicompartmental systems: Modeling and implications. In *International Conference on Neural Information Processing*, pp. 523–531, Springer, 2015.

[27] O.A. Moldes, J.C. Mejuto, R. Rial-Otero and J. Simal-Gandara: A critical review on the applications of artificial neural networks in winemaking technology. *Critical Reviews in Food Science and Nutrition*, **57**(13), (2017), 2896–2908.

[28] E. Oja and S. Kaski: *Kohonen Maps*. Elsevier Science, 1999.

[29] M. O'Shea, R. Colbert, L. Williams and S. Dunn: Nitric oxide compartments in the mushroom bodies of the locust brain. *Neuroreport*, **9**(2), (1998), 333–336.

[30] J.H. Park, V.A. Straub and M. O'Shea: Anterograde signaling by nitric oxide: Characterization and in vitro reconstitution of an identified nitrergic synapse. *Journal of Neuroscience*, **18**(14), (1998), 5463–5476.

[31] P. Perner: *Machine Learning and Data Mining in Pattern Recognition: 11th International Conference*, MLDM 2015, Hamburg, Germany, July 20-21, 2015, Proceedings. Lecture Notes in Computer Science, Springer International Publishing, 2015.

[32] A. Philippides, P. Husbands and M. O'Shea: Four-dimensional neuronal signaling by nitric oxide: a computational analysis. *Journal of Neuroscience*, **20**(3), (2000), 1199–1207.

[33] C. Pozrikidis: *The Fractional Laplacian*. CRC Press, 2016.

[34] G. Pölzlbauer: Survey and comparison of quality measures for self-organizing maps.

[35] M. Senapati: *Advanced Engineering Chemistry*. Laxmi Publications, 2006.

[36] S.H. Snyder and D.S. Bredt: Nitric oxide as a neuronal messenger. *Trends in Pharmacological Sciences*, **12** (1991), 125–128.

[37] J.J. Thomson: On the structure of the atom: an investigation of the stability and periods of oscillation of a number of corpuscles arranged at equal intervals around the circumference of a circle; with application of the results to the theory of atomic structure. *Philosophical Magazine Series 6*, **7**(39), (1904), 237–265.

[38] A. Urmos, Z. Farkas, M. Farkas, T. Sandor, L.T. Koczy and A. Nemcsics: Fuzzy and kohonen som based classification of different 0d nanostructures. In *2017 IEEE 15th International Symposium on Applied Machine Intelligence and Informatics (SAMI)*, pp. 365–370, 2017.

[39] L. Wood and J. Garthwaite: Models of the diffusional spread of nitric oxide: implications for neural nitric oxide signalling and its pharmacological properties. *Neuropharmacology*, **33**(11), (1994), 1235–1244.

[40] Y. Yun: The moments of a diffusion process. *Statistics & Probability Letters*, **138** (2018), 36–41.