Slovak University of Technology in Bratislava
Faculty of Mechanical Engineering
17th Conference on Applied Mathematics **APLIMAT 2018**

Proceedings

GROUPS AND HYPERGROUPS OF ARTIFICIAL NEURONS

CHVALINA Jan (CZ), SMETANA Bedřich (CZ)

Abstract. When we study structure of the most used artificial neural network - multilayer perceptron and functionality of artificial neuron, there is possibility using several ways to describe function and neural network properties on the basis of known algebraic structures, vector spaces and graphs theory or properties of relations. Using certain analogy with relations between descriptions of differential equations certain quality there is developed access to new view point on these subjects. In this paper some concepts of description and modelling systems of neurons are investigated.

Keywords: Neural network, transposition hypergroups, linear ordinary differential operators, groups of neurons

Mathematics subject classification: Primary 92 B 20; Secondary 20 N 20, 43 A 62, 47 E 05

1 Introduction

The beginning of the establishment of neural networks is considered to work Warren McCulloch and Walter Pitts of 1943, which created a very simple mathematical model of a neuron, which is the basic cell of the nervous system [3, 13, 14, 25]. The numerical values of the parameter in this model were predominantly bipolar, i.e. from the set $\{-1, 0, 1\}$. They showed that the simplest types of neural networks can in principle compute any arithmetic or logic function.

A neuron called also as artificial or formal neuron is the basic stone of the mathematical model of any neural network. Its design and functionalites are derived from observation of a biological neuron that is basic building block of biological neural networks (systems) which includes the brain, spinal cord and peripheral ganglia. In case of artificial neuron the information comes into the body of an artificial neuron via inputs that are weighted (i.e. each input can be individualy multiplied with a weight). The body of an artificial neuron then sums the weighted inputs, bias and "processes" the sum with a transfer function. At the end an artificial neuron passes the processed information via outputs.

Neuron activity can be described mathematically: Capturing signal and transmission in neurons, there is created potential P:

$$P = w_1 * x_1 + w_2 * x_2 + ... + x_n * w_n n$$

If the potential is sufficiently large, the neuron transmits a signal y:

$$y = 1, if\ P > w_0,\ otherwise\ y = 0.$$

The condition that $P \ > w_0$ can be overridden by activating function $f(P)$. The entire activity of neurons can then enroll in one mathematical relationship where w_0 is a negative number that represents the threshold that shall overcome potential. Formally, the transfer function can have a zero threshold and a neural boundary with the negative sign being understood as the weight, the so-called bias $w_0 = \ ^{-}\theta$ of another formal input $x_0 = 1$ with a constant unit value. The value of the internal potential y_in,where

$$\mathbf{y\_in} = \sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{w_i x_i}.$$

after reaching the value $\mathbf{b}(\mathbf{w0} = \mathbf{b} - \mathbf{bias})$ it invokes the output state **y** of the neuron **Y**, axon pulse. Increasing the output values $y = y\_in$ when the potential value of **b** is given by the activating (transfer) function f. In general, a single-layer neural network is not capable of solving all tasks. Therefore the most commonly used type is a multilayer feedforward network with a backpropagation learning method. From paper focus is interesting the linear transfer function. The output of a linear transfer function is equal to its input:

$$a = n,$$

as illustrated in Figure.

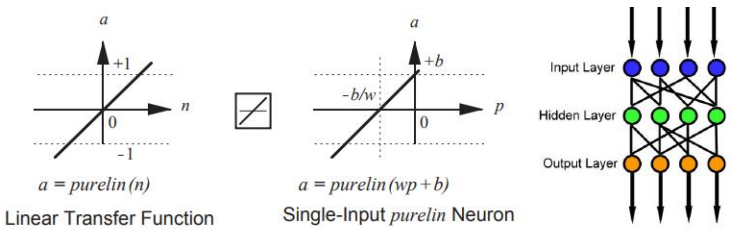

Fig. 1. Feedforward neural network and linear transfer function.

Neurons with this transfer function are used in the ADALINE networks. The output neuron expression can be written in matrix form:

$$\mathbf{y} = \mathbf{W} * \mathbf{x} + \mathbf{b},$$

where the matrix for the single neuron case has only one row. Output and input product of artificial neurons can be the same way interpreted as vectors of input or output linear vector spaces. Now the neuron output of multilayer neural network can be written according terms of matrix form description as:

$$\mathbf{y^n} = \mathbf{f^n}(\mathbf{W^n f^{n-1}}(\mathbf{W^{n-1} f^{n-2}} ...(\mathbf{W^1 x} + \mathbf{b^1}) + \mathbf{b^2}) ... + \mathbf{b^n})$$

for n-layer neural network. For further development is necessary describe basic terms from usage linear ordinary differential operators.

Artificial neural networks can be viewed as a weghted directed graphs in which artificial neurons are nodes and directed edges with weight are connections between neuron outputs and neuron inputs.

Recall that in the framework of Artificial neural networks perceptrons is a network of simple neurons called percetrons. The basic concept of a single perceptron was introduced by Rosenblatt in the year 1958. The perceptron computes a single output from multiple real-valued inputs by forming a linear combination according to its input weights and then possibly putting the output through some nonlinear activation function. As usually mathematically this can be written as:

$$y = \varphi\left(\sum_{i=1}^{n} w_i x_i + b\right) = \varphi(\vec{w}^T \vec{x} + b),$$

where $\vec{w} = (w_1, \ldots, w_n)$ denotes the vector of weights, $\vec{x} = (x_1, \ldots, x_n)$ is the vector of inputs, b is the bias and φ is the activation function.

2 Groups and hypergroups of linear differential operators

Let us mention from the history of sixties and seventies from the past century when Otakar Borůvka and his collaborators and successors begun with the investigation of differential equations using the algebraic and geometrical approach.

The substanding representative of the mentioned school František Neuman wrote in his paper [15]: "Algebraic, topological and geometrical tools together with the methods of the theory of dynamical systems and functional equations make possible to deal with problems concerning global properties of solutions by contrast to the previous local investigations and isolated results." Influence of mentioned ideas is a certain motivating factor of our investigations.

So, we consider linear ordinary differential operators of the form:

$$L_n = \sum_{k=0}^{n} p_k(x) D^k,$$

where $D_k = \frac{d^k}{dx^k}$, $p_k(x)$ is a continuous function on some open interval $J \subset \mathbb{R}$, $k = 0, 1, \ldots, n-1$, $p_n(x) \equiv 1$, i.e. $L_n(y) = 0$ which is a linear homogenous ordinary differential equation of the form:

$$y^(n)(x) + \sum_{k=0}^{n-1} p_k(x) y^{(k)}(x) = 0.$$

By an ordered group we mean (as usually) a triad $(G, \cdot, \leq)$, where $(G, \cdot)$ is a group and $\leq$ is a reflexive, symmetrical and transitive binary relation on the set G such that for any triad $x, y, z \in G$ with the property $x \leq y$ also $x \cdot z \leq y \cdot z$, $z \cdot x \leq z \cdot y$ is satisfied. Further, $[a)_{\leq} = \{x \in G; a \leq x\}$ is the principal end generated by $a \in G$. To any element $a \in G$ there is assigned a pair of mappings $\lambda_a : G \to G, \rho_a : G \to G$, which are called a left translation, a right translation, respectively, determined by the element $a \in G$, i.e. $\lambda_a(x) = a \cdot x$, $\rho_a(x) = x \cdot a$.(Of course, in the case of a commutative group $(G, \cdot)$ we have $\lambda_a \equiv \rho_a$). Notice, that a group with an ordering $(G, \cdot, \leq)$ is an ordered group if and only if all its left and right translations $\lambda_a, \rho_a, a \in G$ are order-preserving mappings, i.e. isotone selfmaps of the ordered set $(G, \leq)$.

The following lemma which is crutial for further constructions is proved in [5, 7] (the Czech version is proved in [6], pp. 146, 147).

Lemma 1. *Let $(G, \cdot, \leq)$ be an ordered group. Define a hyperoperation $* : G \times G \to \mathcal{P}(G)^*$ by*

$$a * b = [a, b)_{\leq} (= \{x \in G; a \cdot b \leq x\})$$

for all pairs of elements $a, b \in G$. Then $(G, *)$ is a hypergroup which is commutative if and only if the group $(G, \cdot)$ is commutative.

Application of the above lemma and many new results are obtained in papers of Michal Novák. See at least titles [20, 21, 22].

For present following results we use similar notation published in [7]. So there $\mathbb{R}$ stands for the set of all reals, $J \subset \mathbb{R}$ is an open interval (bounded or unbounded) of real numbers, $\mathbb{C}^k(J)$ is the ring (with respect to usual addition and multiplication of functions) of all real functions with continuous derivatives up to the order $k \geq 0$ including. We write $\mathbb{C}(J)$ instead of $\mathbb{C}^0(J)$. For a positive integer $n \geq 2$ we denote by $\mathbb{A}_n$ the set of all linear homogeneous differential equations of the *n*-th order with continuous real coefficients defined on *J,* i.e.

$$y^{(n)} + p_{n-1}(x)y^{(n-1)} + \cdots + p_0(x)y = 0,$$

(cf. [5, 7, 14, 15, 19]), where $p_k \in \mathbb{C}(J)$, $k = 0, 1, \cdots, n-1$, $p_0(x) > 0$ for any $x \in J$ (this is not essential restriction). Denote $L(p_0, \cdots, p_{n-1}) : \mathbb{C}^n(J) \to \mathbb{C}^n(J)$ the above defined linear operator defined by

$$L(p_0, \cdots, p_{n-1})(y) = y^{(n)} + p_{n-1}(x)y^{(n-1)} + \cdots + p_0(x)y$$

and put

$$\mathbb{LA}_n(J) = \{L(p_0, \cdots, p_{n-1}); p_k \in \mathbb{C}(J),\, p_0 > 0\}.$$

Further $\mathbb{N}_0(n) = \{0, 1, \cdots, n-1\}$ and δ_{ij} stands for the Kronecker δ, $\overline{\delta_{ij}} = 1 - \delta_{ij}$. For any $m \in \mathbb{N}_0(n)$ we denote by $\mathbb{LA}_n(J)_m$ the set of all linear differential operators of the $n-$th order $L_0(p_0, \cdots, p_{n-1}) : \mathbb{C}^n(J) \to \mathbb{C}(J)$, where $p_k \in \mathbb{C}(J)$ for any $k \in \mathbb{N}_0(n)$, $p_m \in \mathbb{C}_+(J)$, (i.e. $p_m(x) > 0$ for each $x \in J$). Using the vector notation $\vec{p}(x) = (p_0(x), \cdots, p_{n-1}(x))$, $x \in J$ we can write $L_n(\vec{p}_0)y = y^{(n)} + (\vec{p}(x), (y, y', \cdots, y^{(n-1)}))$, (i.e. a scalar product).

We define a binary operation "$\circ_m$" and a binary relation "$\leq_m$ on the set $\mathbb{LA}_n(J)_m$ in this way:

For arbitrary pair $L(\vec{p}), L(\vec{q}) \in \mathbb{LA}_n(J)_m$, $\vec{p} = (p_0, \cdots, p_{n-1})$, $\vec{q} = (q_0, \cdots, q_{n-1})$ we put $L(\vec{p}) \circ_m L(\vec{q}) = L(\vec{u})$, $\vec{u} = (u_0, \cdots, u_{n-1})$, where

$$u_k(x) = p_m(x)q_k(x) + (1 - \delta_{km})p_k(x),\, x \in J$$

and $L(\vec{p}) \leq L(\vec{q})$ whenewer $p_k(x) \leq q_k(x)$, $k \in \mathbb{N}_0(n)$, $p_m(x) = q_m(x)$, $x \in J$. Evidently, $(\mathbb{LA}_n(J)_m, \leq_m)$ is an ordered set.

In paper [7] there is presented the sketch of the proof of the following lemma:

Lemma 2. *The triad* $(\mathbb{LA}_n(J)_m, \circ_m, \leq_m)$ *is an ordered (noncommutative) group.*

In what follows we will construct a group and hypergroup of artificial neurons using the above mentioned approach.

3 Groups and hypergroups of artificial neurons

As it is mentioned in the dissertation [3] neurons are the atoms of neural computation. Out of those simple computational units all neural networks are build up. The output computed by a neuron can be expressed using two functions $y = g(f(w, x))$. The details of computation consist in several steps: In a first step the input to the neuron, $x := \{x_i\}$, is associated with the weights of the neuron, $w := \{w_i\}$, by involving the so-called propagation function f. This can be thought as computing the activation potential from the pre-synaptic activities. Then from that result the so-called activation function

g computes the output of the neuron. The weights, which mimic synaptic strenght, constitute the adjustable internal parameters of the neuron. The process of adapting the weights is named learning.

From the biological point of view it is advisible to use an integrative propagation function. And therefore convenient choice would be to use the weighted sum of the input $f(w,x) = \sum_i w_i x_i$, that is the activation potential equals to the scalar product of input and weights. In fact, the most popular propagation function since the dawn of neural computation, however it is ofen used in the slightly different form:

$$f(w,x) = \sum_i w_i x_i + \Theta, \qquad (*)$$

The special weight Θ is called bias. Applying $\Theta(x) = 1$ for $x > 0$ and $\Theta(x) = 0$ for $x < 0$ as the above activation function yields the famous perceptron of Rosenblatt. In that case the function Θ works as a threshold.

Besides $(*)$ there are, of course, many other possible propagation functions. If $(*)$ is supplemented with the identity as activation function and real-valued domains are given a real linear neuron $\mathsf{y} = \sum_i w_i x_i + \Theta$ is obtained. This real linear neuron can be seen as an example of a Clifford neuron. Denoting by $\mathfrak{C}_{p,q,r}$ the unique universal Clifford algebra corresponding to a standard quadratic space $(\mathbb{R}^{p+q+r}, Q)$, $Q(x) = x^2$ and by $\otimes_{p,q,r}$ the geometric product of the algebra $\mathfrak{C}_{p,q,r}$ (cf. [3] ch. 2, part 2.1) we can say that a Clifford Neuron(CN) computes the following function from $(\mathfrak{C}_{p,q,r})^n$ to $\mathfrak{C}_{p,q,r}$:

$$\mathsf{y} = \sum_{i=1}^{n} w_i \otimes_{p,q,r} x_i + \Theta.$$

([3], Definition 3.1).

It is to be noted that deep neural networks contain multiple non-linear hidden layers and this makes them very expressive models that can learn very complicated relationships between their inputs and outputs.

To estimate networks parameters Hinton et all [25] proposed the wake-sleep algorithm. As models serve $q_i = \sigma\left(\sum_j s_i \Phi_{ij} + \Phi_{0j}\right)$ for the position "wake"
and $p_j = \sigma\left(\sum_k s_k \Theta_{kj} + \Theta_{0j}\right)$ for the position "sleep". Similar probability functions occure within the model description of the socalled Restricted Boltzmann Machines - [25].

The formal description of the architecture of the general Neural Abstractions Pyramid contains formally similar function mentioned above, given by the mathematical model of an artificial neuron. The basic processing element consists of the P_{kl} projection units and a single output unit. The activity $a^t_{ijkl} \in \mathbb{R}$ of the feature cell at position (i,j) for feature array k in layer l at time t is computed as follows:

$$a^t_{ijkl} = \psi_{kl}\left(\sum_{p=1}^{P_{kl}} v^p_{kl} b^{l_p}_{ijkl} + v^0_{kl}\right).$$

The output unit computes a weighted sum of the projection potentials $b^{tp}_{ijkl} \in \mathbb{R}$ with the weighting factors described by $v^p_{kl} \in \mathbb{R}$. A bias value of v^0_{kl} is also added to the sum before it is passed through the output transfer function ψ_{kl}.

The computation of the individual projection potentials is described by:

$$b^{tp}_{ijkl} = \varphi_{kl}\left(\sum_{q=1}^{Q^p_{kl}} w^{pq}_{kl} a^{t'}_{i'j'k'l'} + w^{p0}_{kl}\right).$$

The processing element that computes a feature cells consists of P_{kl} projection units and output unit that produces the activity a^t_{ijkl}. The output unit computes the weighted sum of the potentials b^{tp}_{ijkl} of the individual projections and passes this sum through a transfer function ψ_{kl}. Each projection unit computes the weighted sum of activities $a^{l'}_{i'j'k'l'}$ with the weighting factors described by $w^{pq}_{kl} \in \mathbb{R}$. The number of contributions to a projection p is Q^p_{kl}. In addition, a bias value of w^{p0}_{kl} is added before the sum is passed trough the projection transfer function ϕ^p_{kl}.

So, recall the well-known mathematical description of a formal neuron:

Let $F : \mathbb{R} \to \mathbb{R}$ be a general non-linear (or piece-wise linear) transfer function. Then the action of a neuron can be expressed as this model:

$$y(k) = F\left(\sum_{i=1}^{m} w_i(k)x_i(k) + b\right),$$

where $x_i(k)$ is input value in discrete time k where i goes from 0 to m, $w_i(k)$ is weight value in discrete time where i goes from 0 to m, b is bias, $y_i(k)$ is output value in discrete time k.

Notice that in some very special cases the transfer function F can be also linear. Transfer function defines the properties of artificial neuron and can be any mathematical function. Usually it is chosen on the basis of problem that artificial neuron(artificial neural network) needs to solve and in most cases it is taken(as mentioned above) from the following set of functions: step function, linear function and non-linear(sigmoid) function.

In what follows we will consider a certain generalization of classical artificial neurons mentioned above consisting in such a way that inputs x_i and weight w_i will be functions of an argument t belonging into a linearly ordered (tempus) set T with the least element 0. As the index set we use the set $\mathbb{C}(J)$ of all continuous functions defined on an open interval $J \subset \mathbb{R}$. So, denote by W the set of all non-negative functions $w : T \to \mathbb{R}$ forming a subsemiring of the ring of all real functions of one real variable $x : \mathbb{R} \to \mathbb{R}$. Denote by $Ne(\vec{w}_r) = Ne(w_{r1}, \ldots, w_{rn})$ for $r \in \mathbb{C}(J)$, $n \in \mathbb{N}$ the mapping

$$y_r(t) = \sum_{k=1}^{n} w_{r,k}(t)x_{r,k}(t) + b_r$$

which will be called the artificial neuron with the bias $b_r \in \mathbb{R}$. By $\mathbb{AN}(T)$ we denote the collection of all such artificial neurons.

Neurons are usually denoted by capital letters X, Y or X_i, Y_i, nevertheless we use also notion $Ne(\vec{w})$, where $\vec{w} = (w_1, \ldots, w_n)$ is the vector of weights.

We suppose - for the sake of simplicity - that transfer functions (activation functions) φ, $\sigma(or\ f)$ are the same for all neurons from the collection $\mathbb{AN}(T)$ or the role of this function plays the identity function $f(y) = y$.

Now, similarly as in the case of the collection of linear differential operators above, we will construct a group and hypergroup of artificial neurons.

Denote by δ_{ij} the so called Kronecker delta, $i, j \in \mathbb{N}$, i.e. $\delta_{ii} = \delta_{jj} = 1$ and $\delta_{ij} = 0$, whenever $i \neq j$.

Suppose $Ne(\vec{w}_r),\ Ne(\vec{w}_s) \in \mathbb{AN}(T),\ r, s \in \mathbb{C}(J),\ \vec{w}_r = (w_{r1}, \ldots, w_{r,n}),\ \vec{w}_s = (w_{s1}, \ldots, w_{s,n})$, $n \in \mathbb{N}$. Let $m \in \mathbb{N},\ 1 \leq m \leq n$ be a such an integer that $w_{r,m} > 0$. We define

$$Ne(\vec{w}_r) \cdot_m Ne(\vec{w}_s) = Ne(\vec{w}_u),$$

where

$$\vec{w}_u = (w_{u,1}, \ldots, w_{u,n}) = (w_{u,1}(t), \ldots, w_{u,n}(t)),$$
$$\vec{w}_{u,k}(t) = w_{r,m}(t) w_{s,k}(t) + (1 - \delta_{m,k}) w_{r,k}(t), t \in T$$

and, of course, the neuron $Ne(\vec{w}_u)$ is defined as the mapping $y_u(t) = \sum\limits_{k=1}^{n} w_k(t)x_k(t) + b_u,\ t \in T,\ b_u = b_r b_{s.}$ Further for a pair $Ne(\vec{w}_r),\ Ne(\vec{w}_s)$ of neurons from $\mathbb{AN}(T)$ we put $Ne(\vec{w}_r) \leq_m Ne(\vec{w}_s),\ w_r = (w_{r,1}(t), \ldots, w_{r,n}(t)),\ w_s = (w_{s,1}(t), \ldots, w_{s,n}(t))$ if $w_{r,k}(t) \leq w_{s,k}(t),\ k \in \mathbb{N},\ k \neq m$ and $w_{r,m}(t) = w_{s,m}(t),\ t \in T$ and with the same bias. Evidently $(\mathbb{AN}(T), \leq_m)$ is an ordered set. A relationship (compatibility) of the binary operation "$\cdot_m$" and the ordering $\leq_m$ on $\mathbb{AN}(T)$ is given by this assertion analogical to the above one.

Lemma 3. *The triad* $(\mathbb{AN}(T), \cdot_m, \leq_m)$ *(algebraic structure with an ordering) is a non-commutative ordered group.*

Sketch of the proof:

1. The operation "$\cdot_m$" is evedently asociative.

2. Let $\vec{w}(t) = (w_1(t), \ldots, w_n(t))$, where $w_k(t) = \delta_{km}$ for any $t \in T$. Then the neuron $Ne(\vec{w})$ determined by $y(t) = \sum\limits_{k=1}^{n} w_k(t)x_k(t) + 1$ is the neutral element of the semigroup $(\mathbb{AN}(T),\ \cdot_m)$, i.e. the unit with respect to the binary operation "$\cdot_m$".

3. Inverse elements: Define

$$\bar{w}_k(t) = (w_m(t))^{-1}(w_k(t) + 1)\delta_{km} - w_k(t),$$

   $t \in T$ and any $k = 1, 2, \ldots, n$. Then the inverse element to the neuron $Ne(\vec{w})$, with $y(t) = \sum\limits_{k=1}^{n} w_k(t)x_k(t) + \mathfrak{b}$ within the semigroup $(\mathbb{AN}(T),\ \cdot_m)$ is the neuron $Ne(\vec{\bar{w}})$, determined by $\bar{y}(t) = \sum\limits_{k=1}^{n} \bar{w}_k(t)\bar{x}_k(t) + \mathfrak{b}^{-1}$.

4. Compatibility of the ordering relation $\leq_m$ with the binary operation "$\cdot_m$" (the substitution property):

   Suppose $Ne(\vec{w}_r),\ Ne(\vec{w}_s),\ Ne(\vec{w}_u) \in \mathbb{AN}(T)$ are neurons such that $Ne(\vec{w}_r) \leq_m Ne(\vec{w}_s)$, i.e. $w_{r,m}(t) \equiv w_{s,m}(t),\ w_{r,k}(t) \leq w_{s,k}(t)$ for any index $k \in \{1, 2, \ldots, n\}$, $k \neq m,\ t \in T$. Denote $Ne(\vec{w}_a) = Ne(\vec{w}_r) \cdot_m Ne(\vec{w}_u),\ Ne(\vec{w}_b) = Ne(\vec{w}_s) \cdot_m Ne(\vec{w}_u)$, where $\vec{w}_a(t) = (w_{a,1}(t), \ldots, w_{a,n}(t)),\ \vec{w}_b(t) = (w_{b,1}(t), \ldots, w_{b,n}(t)),\ t \in T$.

   For any index $k \in \{1, 2, \ldots, n\}$ we have

$$w_{a,k(t)} = w_{r,m}(t) w_{u,k}(t) + (1 - \delta_{km}) w_{r,k}(t),$$

$$w_{b,k}(t) = w_{s,m}(t)w_{u,k}(t) + (1 - \delta_{km})w_{s,k}(t),\ t \in T,$$

Since $w_{r,m}(t) \equiv w_{s,m}(t),\ w_{r,k}(t) \le w_{s,k}(t),\ k \ne m,\ t \in T$, for $k = m$ there holds:

$$w_{a,m}(t) = w_{r,m}(t)w_{u,k}(t) = w_{s,m}(t)w_{u,m}(t),\ t \in T$$

and for $k \ne m$ we have

$$w_{a,k}(t) = w_{r,m}(t)w_{u,k}(t) + w_{r,k}(t) \le w_{s,m}(t)w_{u,k}(t) + w_{s,k}(t) = w_{b,k}(t),\ t \in T,$$

thus

$$Ne(\vec{w}_r) \cdot_m Ne(\vec{w}_u) \le_m Ne(\vec{w}_s) \cdot_m Ne(\vec{w}_u).$$

Similarly we obtain that

$$Ne(\vec{w}_u) \cdot_m Ne(\vec{w}_r) \le_m Ne(\vec{w}_u) \cdot_m Ne(\vec{w}_s).$$

Denoting

$$\mathbb{AN}_1(T)_m = \{Ne(\vec{w}); \vec{w} = (w_1, \dots, w_n),\ w_k \in \mathbb{C}(T),\ k = 1, \dots, n,\ w_m(t) \equiv 1\},$$

we get the following assertion:

Proposition 1. *Let $T = \langle 0, t_0) \subset \mathbb{R},\ t_0 \in \mathbb{R} \cup \{\infty\}$. Then for any positive integer $n \in \mathbb{N},\ n \ge 2$ and for any integer m such that $1 \le m \le n$ the semigroup $(\mathbb{AN}_1(T)_m,\ \cdot_m)$ is an invariant subgroup of the group$(\mathbb{AN}(T)_m,\ \cdot_m)$.*

1. Suppose $Ne(\vec{w}_r),\ Ne(\vec{w}_s) \in (\mathbb{AN}_1(T)_m$ are arbitrary neurons. Then

$$\vec{w}_r = (w_{r,1}(t), \dots, w_{r,m}(t), \dots, w_{r,n}(t)),$$
$$\vec{w}_s = (w_{s,1}(t), \dots, w_{s,m}(t), \dots, w_{s,n}(t)),$$

where $w_{r,m}(t) \equiv 1,\ w_{s,m}(t) \equiv 1$. Evidently the formal neuron $Ne(\vec{w})$, where $\vec{w}(t) = (w_1(t), \dots, w_n(t))$ with $w_k(t) = \delta_{km},\ t \in T$, acting by

$$y(t) = \sum_{k=1}^{n} w_k(t)x_k(t) + 1,$$

which is, of course, neutral element of the group $(\mathbb{AN}(T)_m,\ \cdot_m)$ also belongs to $\mathbb{AN}_1(T)_m$. Denoting $Ne(\vec{w}_v) = Ne(\vec{w}_r). Ne(w_s),\ \vec{w}_v = (w_{v,1}, \dots, w_{v,n})$ we have

$$w_{v,m}(t) = w_{r,m}(t)(w_{s,m}(t))^{-1} = 1$$

for any $t \in T$, thus $Ne(\vec{w}_v) \in \mathbb{AN}_1(T)_m$, consequently we obtain that $(\mathbb{AN}_1(T)_m,\ \cdot_m)$ is a subgroup of the group $(\mathbb{AN}(T)_m,\ \cdot_m)$.

2. Now suppose $Ne(\vec{w}_r) \in \mathbb{AN}(T)_m,\ Ne(\vec{w}_s) \in \mathbb{AN}_1(T)_m$, where $\vec{w}_r(t),\ \vec{w}_s(t)$ are vectors of function as above. Denoting

$$Ne(\vec{w}_u) = Ne^{-1}(\vec{w}_r) \cdot_m Ne(\vec{w}_s) \cdot_m Ne(\vec{w}_r),$$

where $\vec{w}_u(t) = \vec{w}(t) = (w_{u,1}(t), \dots, w_{u,n}(t)),\ t \in T$, then $\vec{w}_{u,m}(t) = (\vec{w}_{r,m}(t))^{-1} \cdot_m \vec{w}_{s,m}(t) \cdot_m \vec{w}_{r,m}(t) = \vec{w}_{s,m}(t) = 1$ for any $t \in T$, thus $Ne(\vec{w}_u \in \mathbb{AN}_1(T)_m$, which means that

$$Ne^{-1}(\vec{w}_s) \cdot_m \mathbb{AN}_1(T)_m \cdot_m Ne(\vec{w}_s) \subset \mathbb{AN}_1(T)_m,$$

therefore the group $(\mathbb{AN}_1(T)_m,\ \cdot_m)$ is an invariant subgroup of the group $(\mathbb{AN}(T)_m,\ \cdot_m)$.

If $m, n \in \mathbb{N}$, $1 \le m \le n-1$, then a certain relationship between groups $(\mathbb{AN}_r(T)_m, \cdot_m)$, $(\mathbb{LA}(T)_{m+1}, \circ_{m+1})$ is contained in the following proposition:

Proposition 2. *Let $t_0 \in \mathbb{R}$, $t_0 > 0$, $T = \langle 0, t_0) \subset \mathbb{R}$ and $m, \, ,n \in \mathbb{N}$ are integers such that $1 \le m \le n-1$. Define a mapping $F : \mathbb{AN}_n(T)_m \to \mathbb{LA}_n(T)_{m+1}$ by this rule: For an arbitrary neuron $Ne(\vec{w}_r \in \mathbb{AN}_n(T)_m$, where $\vec{w}_r = (w_{r,1}(t), \ldots, w_{r,n}(t)) \in [\mathbb{C}(T)]^n$ we put $F(Ne(\vec{w}_r)\,) = L(w_{r,1}, \ldots, w_{r,n}) \in \mathbb{LA}_n(T)_{m+1}$ with the action :*

$$L(w_{r,1}, \ldots, w_{r,n})y(t) = \frac{d^n y(t)}{dt^n} + \sum_{k=1}^{n} w_{r,k}(t) \frac{d^{k-1}(t)}{dt^{k-1}},\ y \in \mathbb{C}^n(T).$$

Then the mapping $F : \mathbb{AN}_n(T)_m \to \mathbb{LA}_n(T)_{m+1}$ is a homomorphism of the group $(\mathbb{AN}_n(T)_m, \cdot_{m)}$ into the group $(\mathbb{LA}_n(T)_{m+1}, \circ_{m+1})$.

Consider $Ne(\vec{w}_r)$, $Ne(\vec{w}_s) \in \mathbb{AN}_n(T)_m$ and denote $F(Ne(\vec{w}_r)) = L(w_{r,1}, \ldots, w_{r,n})$, $F(Ne(\vec{w}_s = L(w_{s,1}, \ldots, w_{s,n})$. Denote $Ne(\vec{w}_u) = Ne(\vec{w}_r) \cdot_m Ne(\vec{w}_s)$. There holds

$$F(Ne(\vec{w}_r) \cdot_m Ne(\vec{w}_s)) = F(Ne(\vec{w}_u)) = L(w_{u,1}, \ldots, w_{u,n}),$$

where

$$L(w_{u,1}, \ldots, w_{u,n})y(t) = y^{(n)}(t) + \sum_{k=1}^{n} w_{u,k}(t) y^{(k-1)}(t).$$

Here $w_{u,k}(t) = w_{r,m+1}(t) w_{s,k}(t) + w_{r,k}(t),\ k \ne m$, and $w_{u,m+1}(t) = w_{r,m+1}(t) w_{s,m+1}(t)$.
Then $L(w_{u,1}, \ldots, w_{u,n}) = L(w_{r,1}, \ldots, w_{r,n}) \cdot_m L(w_{s,1}, \ldots, w_{s,n}) = F(Ne(\vec{w}_r)) \cdot_m F(Ne(\vec{w}_s))$. The neutral element $Ne(\vec{w}) \in \mathbb{AN}_n(T)_m$ is also mapped onto the neutral element of the group $(\mathbb{L}_n\mathbb{A}(T)_{m+1}, \cdot_{m+1})$, thus the mapping $F : (\mathbb{AN}_n(T)_m, \cdot_m) \to (\mathbb{L}_n\mathbb{A}(T)_{m+1}, \circ_{m+1})$ is a group homomorphism.

Now, using the construction described in the above Lemma we obtain the final transpozition hypergroup (called also non-commutative join space). Denote by $\mathbb{P}(\mathbb{AN}(T)_m)^*$ the power set of $\mathbb{AN}(T)_m$ consisting of all nonempty subsets of the last set and define a binary hyperoperation

$$*_m : \mathbb{AN}(T)_m \times \mathbb{AN}(T)_m \to \mathbb{P}(\mathbb{AN}(T)_m)^*$$

by the rule

$$Ne(\vec{w}_r) *_m Ne(\vec{w}_s) = \{Ne(\vec{w}_u);\ Ne(\vec{w}_r) \cdot_m Ne(\vec{w}_s) \le_m Ne(\vec{w}_u)\}$$

for all pairs $Ne(\vec{w}_r)$, $Ne(\vec{w}_s) \in \mathbb{AN}(T)_m$. More in detail if $\vec{w}(u) = (w_{u,1}, \ldots, w_{u,n})$, $\vec{w}(r) = (w_{r,1}, \ldots, w_{r,n})$, $\vec{w}(s) = (w_{s,1}, \ldots, w_{s,n})$, then $w_{r,m}(t) w_{s,m}(t) = w_{u,m}(t)$, $w_{r,m}(t) w_{s,k}(t) + w_{r,k}(t) \le w_{u,k}(t)$, if $k \ne m$, $t \in T$. Then we have that $(\mathbb{AN}(T)_m, *_m)$ is a non-commutative hypergroup. The above defined invariant (termed also normal) subgroup $(\mathbb{AN}_1(T)_m, \cdot_m)$ of the group $(\mathbb{AN}(T)_m, \cdot_m)$ is the carried set of a subhypergroup of the hypergroup $(\mathbb{AN}(T)_m, *_m)$ and it has certain significant properties.

Using certain generalization of methods from [7] we obtain after investigation of constructed structures this result:

Let $T = \langle 0, t_0) \subset \mathbb{R}$, $t_0 \in \mathbb{R} \cup \{\infty\}$. Then for any positive integer $n \in \mathbb{N}$, $n \ge 2$ and for any integer m such that $1 \le m \le n$ the hypergroup $(\mathbb{AN}(T)_m, *_m)$,where

$$\mathbb{AN}(T)_m = \{Ne(\vec{w}_r);\ \vec{w}_r = (w_{r,1}(t), \ldots, w_{r,n}(t)) \in [\mathbb{C}(T)]^n,\ w_{r,m}(t) > 0,\ t \in T\},$$

is a transposition hypergroup (i.e. a non-commutative join space) such that $(\mathbb{AN}(T)_m, *_m)$ is its subhypergroup, which is

- invertible (i.e. $Ne(\vec{w}_r)/Ne(\vec{w}_s) \cap \mathbb{AN}_1(T)_m \neq \emptyset$ implies $Ne(\vec{w}_s)/Ne(\vec{w}_r) \cap \mathbb{AN}_1(T)_m \neq \emptyset$ and $Ne(\vec{w}_r)\ Ne(\vec{w}_s) \cap \mathbb{AN}_1(T)_m \neq \emptyset$ implies $Ne(\vec{w}_s)\ Ne(\vec{w}_r) \cap \mathbb{AN}_1(T)_m \neq \emptyset$ for all pairs of neurons $Ne(\vec{w}_r),\ Ne(\vec{w}_s) \in \mathbb{AN}_1(T)_m$,

- closed (i.e. $Ne(\vec{w}_r)/Ne(\vec{w}_s) \subset \mathbb{AN}_1(T)_m$, $Ne(\vec{w}_r) \setminus Ne(\vec{w}_s) \subset \mathbb{AN}_1(T)_m$ for all pairs $Ne(\vec{w}_r), /, Ne(\vec{w}_s) \in \mathbb{AN}_1(T)_m$,

- reflexive (i.e. $Ne(\vec{w}_r)\ \mathbb{AN}_1(T)_m = \mathbb{AN}_1(T)_m/Ne(\vec{w}_r)$ for any neuron $Ne(\vec{w}_r) \in \mathbb{AN}(T)_m$ and

- normal (i.e.$Ne(\vec{w}_r) * \mathbb{AN}_1(T)_m = \mathbb{AN}_1(T)_m * Ne(\vec{w}_r)$ for any neuron $Ne(\vec{w}_r) \in \mathbb{AN}(T)_m$.

Remark A certain generalization of the formal (artificial) neuron can be obtained from expression of linear differential operator of the *n*-th order. Recall the expression of formal neuron with inner potential $y_{-in} = \sum_{k=1}^{n} w_k(t)x_k(t)$, where $\vec{x}(t) = (x_1(t), \ldots, x_n(t))$ is the vector of inputs, $\vec{w}(t) = (w_1(t), \ldots, w_n(t))$ is the vector of weights. Using the bias b of the considered neuron and the transfer function σ we can expressed the output as $y(t) = \sigma\left(\sum_{k=1}^{n} w_k(t)x_k(t) + b\right)$.

Now consider a tribal function $u : J \to \mathbb{R}$, where $J \subseteq \mathbb{R}$ is an open interval; input are derived from the function $u \in \mathbb{C}^n(J)$ as it follows: Inputs $x_1(t) = u(t), x_2 = \frac{du(t)}{dt}, \ldots, x_n(t) = \frac{d^{n-1}(t)}{dt^{n-1}}$, $n \in \mathbb{N}$. Further the bias $b = b_0\frac{d^n u(t)}{dt^n}$. As weights we use the continuous function $w_k : J \to \mathbb{R}$, $k = 1, \ldots, n-1$.

Then formula

$$y(t) = \sigma\left(\sum_{k=1}^{n} w_k(t)\frac{d^{k-1}u(t)}{dt^{k-1}} + b_0\frac{d^n u(t)}{dt^n}\right)$$

is a description of the action of the neuron Dn which will be called a formal(artificial) differential neuron. This approach allows to use solution spaces of corresponding linear differential equations.

4 Conclusion

Neural nets and neural computation form wide topics with interesting history and with many applications in science and number of technical utilizations. In references there is presented a certain, but very restricted choice of publications [2, 3, 12-14, 25-30]. The new chapter in the history of neural computation is attributed with the names Rumelhart and McClelland, knowned as members of PDP group, co-authors of algorithm backpropagation. Around the time neural networks were widely recognized as leading directly forwards real artificial inteligence.Our considerations are based on algebraic approach using classical structures in which are investigated in the present time. These investigations allow further development.

Acknowledgement The second author was supported by a Brno University of Technology grant No. FEKT-S-17-4225.

References

[1] BERÁNEK, J., CHVALINA, J.: *Noncommutative join hypergroups of affine transformations of ordered fields*, In: Univ. South Bohemia, České Budějovice, Depth. Math. Report Series Vol.10 (2002), pp. 15-22.

[2] BEHNKE, S.: *Hierarchical Neural Networks for Image Interpretation*, Lecture Notes in Computer Science, vol. 2766, Springer, Heidelberg (2003), pp. 245.

[3] BUCHHOLZ, S.: *A theory of neural computation with Clifford-algebras*, Technical Report Number 0504, Christian-Albrechts-Universität zu Kiel, Institut für Informatik und Praktische Mathematik (2005), pp.1-135.

[4] CHVALINA, J.: *Commutative hypergroups in the sense of Marty and ordered sets*, Proceedings of the Summer School an General Algebra and Ordered Sets, Olomouc (Czech Republic),1994, pp.19-30.

[5] CHVALINA, J.: *Join spaces of linear ordinary differential operators of the second order*, Colloquium on Differential and Difference Equations, CDDE 2002, Proceedings (M. Bartušek, O. Došlý, ed.) Folia Fac. Sci. Nat. Univ. Masarykianae Brunensis, Mathematica 13, MU Brno 2003, p. 77-86. Masaryk Univ. Brno 1995, 305 pp. (in Czech).

[6] CHVALINA, J.: *Functional Graphs, Quasi-ordered Sets and Commutative Hypergroups*,Masaryk Univ. Brno 1995, pp. 205 (in Czech).

[7] CHVALINA, J., CHVALINOVÁ, L.: *Modelling of join spaces by n- th order linear ordinary differential operators.* In: Fourth International Conference APLIMAT 2005. Bratislava: Slovak University of Technology, Bratislava, 2005. pp. 279-284. ISBN: 80-969264-2- X.

[8] CHVALINA, J., HOŠKOVÁ, Š.: *Modelling of join spaces with compatible proximities by first-order linear partial differential operators.* Italian Journal of Pure and Applied Math. 2004, pp. 177-190.

[9] CORSINI, P.: *Prolegomena of Hypergroup Theory.* Aviani Editore Tricesimo 1993.

[10] CORSINI, P., LEOREANU, V.: *Applications of Hyperstructure Theory.* Kluwer Academic Publishers, Dordrecht/Boston/London 2003.

[11] GOODMANN, G., S., HOFFMANN-JORGENSEN, J.: *Remarks on the calculus of set-valued mappings*, Preprint Series 1968/69 No. 10, Mat. Institut Aarhus Univ., pp. 3.

[12] HAGAN, M., DEMUTH, H., BEALE, M.: *Neural Network Design*, Boston, MA: PWS Publishing, 1996.

[13] KRENKER, A., BEŠTER, J., KOS, A.: *Introduction to the artificial neural networks*. K. Suzuki (Ed.), Artificial neural networks – Methodological advances and biomedical applications, In: Tech, Rijeka (2011), pp. 3-18.

[14] McCULLOCH, W., PITTS, W.: *A logical calculus of the ideas immanent in nervous activity.* Bulletin of Mathematical Biophysics, vol. 5, 1943, pp. 115-133.

[15] NEUMAN, F.: *From local to global investigations of linear differential equations of the n-th order.* Jahrbuch Uberlicke Mathematik 1984, pp. 55-80.

[16] NEUMAN, F.: *Global Properties of Linear Ordinary Differential Equations* Academia Praha - Kluwer Academic Publishers Dordrecht/Boston/London 1991, pp. 320.

[17] NEUMAN, F.: *Global theory of ordinary linear homogenerous differential equations in the real domain.* Aequationes Math. 33 (1987), pp. 123-149.

[18] NEUMAN, F.: *Global theory of ordinary linear homogeneous differential equations in the real domain II.* Aequationes Math. 34 (1987), pp. 1-22.

[19] NEUMAN, F.: *Ordinary linear differential equations - a survey of the global theory.* Equadiff 6, Proc. Internat. Conf. Differential Equations and their Appl., Brno 1985, pp. 59-70.

[20] NOVÁK, M.: *n-ary hyperstructures constructed from binary quasi- ordered semigroups.* Analele Stiintifice Ale Universitatii Ovidius Constanta, Seria Matematica, 2014, roč. 2014 (22), č. 3, pp. 147-168. ISSN: 1224 - 1784.

[21] NOVÁK, M.: *Some basic properties of EL- hyperstructures.* European Journal of Combinatorics, 2013, roč. 2013 (34), č. 2, pp. 446-459. ISSN: 0195 - 6698.

[22] NOVÁK, M.: *On EL - semihypergroups.* European Journal of Combinatorics, 2015, roč. 44 (2015), č. B, Feb. 2015, pp. 274-286. ISSN: 0195 - 6698.

[23] RAČKOVÁ, P.: *Hypergroups of symmetric matrices*, 10th International Congress of Algebraic Hyperstructures and Applications, Proceedings of AHA 2008, University of Defence, Brno (2009), pp. 267-272.

[24] ROSENBERG, I., G.: *Hypergroups and join spaces determined by relations*, Ital. J. Pure Appl. Math 4 (1998), pp. 93 - 101.

[25] SRIVASTAVA, N., HINTON, G., KRIZHEVSKY, A., SUTSKEVER, I., SALAKHUTDINOV, R.: *Dropout: a simple way to prevent neural networks from overfitting.* J. Machine Learning Res. 15, 2014, pp. 1929–1958.

[26] STANĚK, D.: *Rough Sets on State Spaces of Automata.* In Studen EEICT, Proceedings of the 22nd conference. Brno: Vysoké učení technické v Brně, FEKT, 2016. pp. 753-757. ISBN: 978-80-214-5350- 0.

[27] TUČKOVÁ, J.: *Comparison of two Approaches in the Fundamental Frequency Control by Neural Nets.* Proceedings of the 6th Czech-German Workshop in Speech Processing, Praha, September 1996, pp.37.

[28] TUČKOVÁ, J., BOREŠ, P.: *The Neural Network Approach in Fundamental Frequency Control.* Speech Processing: Forum Phoneticum (Ed.: Wodarz,H.-W.), Frankfurt am Main, Hector Verlag. 1997, pp.143-154, ISSN 0341-3144, ISBN 3-930220-10-5.

[29] TUČKOVÁ, J., ŠEBESTA, V.: *Data Mining Approach for Prosody Modelling by ANN in Text-to-Speech Synthesis.* Proc. of the IASTED Inernational Conference on "Artificial Intelligence and Applications - AIA 2001", Ed. M. H. Hamza, ACTA Press Anaheim-Calgary-Zurich, ISBN: 0-88986-301-6, ISSN: 1482-7913, pp. 164-166, September 2001, Marbella, Spain.

[30] VOLNÁ, E.: *Neuronové sítě 1. 2.* vyd. Ostrava: Ostravská univerzita, 2008, pp. 86.

[31] WARNER, M., W.: *Semigroup, group quotient and homogeneous automata*, Inform. Control 47 (1980), pp. 59-66.

Current address

Chvalina Jan, Prof., RNDr., DrSc.
Brno University of Technology, Faculty of Electrical Engineeering and Communication, Department of Mathematics, Technická 8, 616 00 Brno, Czech Republic
Tel. number, e-mail: +420 541146058, chvalina@feec.vutbr.cz

Smetana Bedřich, RNDr.
Brno Univerzity of Defence, Faculty of Military Leadership, Department of Econometrics, Kounicova 65, 662 10 Brno, Czech Republic
Tel. number, e-mail: +420 973443429, bedrich.smetana@unob.cz