Recognition of Generalized Patterns by a Differential Polynomial Neural Network

Ladislav Zjavka
Faculty of Management Science and Informatics
University of Žilina
Žilina, Slovakia
lzjavka@gmail.com

Abstract — A lot of problems involve unknown data relations, identification of which can serve as a generalization of their qualities. Relative values of variables are applied in this case, and not the absolute values, which can better make use of data properties in a wide range of the validity. This resembles more to the functionality of the brain, which seems to generalize relations of variables too, than a common pattern classification. Differential polynomial neural network is a new type of neural network designed by the author, which constructs and approximates an unknown differential equation of dependent variables using special type of root multi-parametric polynomials. It creates fractional partial differential terms, describing mutual derivative changes of some variables, likewise the differential equation does. Particular polynomials catch relations of given combinations of input variables. This type of identification is not based on a whole-pattern similarity, but only to the learned hidden generalized relations of variables.

Keywords - polynomial neural network; dependence of variables identification; differential equation approximation; rational integral function

I. INTRODUCTION

The principal disadvantage of the artificial neural network (ANN) identification in general is the disability of input pattern generalization. ANNs can learn to classify any input patterns but utilize only the absolute values of variables. However, the latter may differ significantly while their relations may be the same. That is why ANNs are able to correctly recognize only similar or incomplete patterns compared to the train set. If the input considered is e.g. a shape moved or sized in the input matrix, the neural network identification will fail. An approach to look at the input vector of variables not as a "pattern" but as a dependent bound point set of N-dimensional space could be attempted. A neural network, which would be able to learn and identify any unknown data relations, is to contain a multi-parametric polynomial functions to catch partial dependence of given inputs. Its response would be the same to all patterns (sets) which variables are performed with the trained dependence, regardless of the actual values [9]. Biological neural cell seems to apply a similar principle. Its dendrites collect electrical signals coming from other neurons. But unlike the artificial neuron, some of the signals already interact in single branches (dendrites) of a neural cell (see Figure 1), likewise the multiplied variables of a multi-parametric polynomial do.

Parameters of polynomial terms can represent the synopsis of the cell dendrites. These weighted combinations are summed in the body cell and transformed into relative values using time-delayed dynamic periodic activation functions (the activated neural cell generates series of time-delayed output pulses, in response to its input signals). Axon passes electrical pulse signals on to dendrites of other neural or effector cells [1]. The period of this function depends on some input variables and seems to represent the derivative part of a partial term of a differential equation composition. Differential polynomial neural network (D-PNN) constructs and tries to approximate an unknown differential equation describing relations of input variables that are not entirely patterns. It forms its output as a generalization of input patterns similar to the ones utilized by the human brain. It creates a structural model of any unknown relationships of input variables description. D-PNN is based on GMDH (Group Method of Data Handling) polynomial neural network, which was created by the Ukrainian scientist Aleksey Ivakhnenko in 1968, when the back-propagation technique was not known yet. He attempted to decompose the complexity of a process into many simpler relationships each described by a low order 2-variable polynomial processing function of a single neuron [2].
II. DIFFERENTIAL POLYNOMIAL NEURAL NETWORK

The basic idea of the D-PNN is to create and approximate a differential equation (DE) (3), which is not known in advance [3], with a special type of root (power) fractional multi-parametric polynomials (5).

\[ a + \sum_{i=1}^{n} b_i \frac{\partial u}{\partial x_i} + \sum_{j=1}^{m} \sum_{i=1}^{n} c_{ij} \frac{\partial^2 u}{\partial x_i \partial x_j} + \ldots = 0 \]

where \( u(x_1, x_2, \ldots, x_d) \) - searched function of all input variables \( a, b_i, b_{ij}, b_{ijk}, b_{ijkl}, \ldots \) - parameters

Fourier’s method of partial DE solution searches the solution in a form of the product of 2 functions, of which at least 1 depends only on 1 variable. A partial derivation of function \( z(x, y) \) of 2 input variables \( x, y \) can be expressed by (4) [4].

\[ \frac{\partial z}{\partial x}(x, y) = f_1(x) \cdot f_2(z) \]

Elementary methods of a differential equation solution express the solution in special elementary functions – polynomials (e.g. Bessel’s functions, Fourier’s power series). Numerical integration of differential equations is based on their approximation through:

- rational integral functions
- trigonometric series

The 1st, and more simple way, has been selected, using the method of integral analogues, which replaces mathematical operators and symbols in DE by the ratio of corresponding operators. Derivatives are replaced by the integral analogues, i.e. derivative operators are removed and simultaneously all operators are replaced by similarly or proportion marks in equations, all vectors are replaced by their absolute values. Dimensional terms are divided by some others, which results in searched non-dimensional likeness criterions [5].

\[ \frac{f_j(x_1, x_2, \ldots, x_n)}{x_i} \]

Each layer of the D-PNN consists of blocks, which contain derivative neurons, one for each fractional polynomial (5), defining the partial derivative dependent change of some input variables. A block also contains additional extended neurons (EN), which form compound functions (15) applying previous layer block outputs. Each block contains a single polynomial (without derivative part), which forms its output entrance into the next hidden layer (Figure 2.). Neurons don’t affect the block output but are applied only for the total output calculation (DE composition). Each neuron has 2 vectors of adjustable parameters \( \mathbf{a}, \mathbf{b} \) and each block contains 1 vector of adjustable parameters of the output polynomial. The root functions of denominators (5) are lower than \( n \), according to the combination degree, which take the polynomials of neurons into competent power degree. They can be replaced by power functions of denominators. Inputs of constant combination degree \((n=2, 3, \ldots)\) forming particular combinations of variables, enter each block, where they are substituted into polynomials (Figure 3.). It is necessary to adjust not only the polynomial parameters, but also the D-PNN’s structure. This means some neurons in terms of role of the DE are to be left out.

\[ \frac{\partial^m f(x_1, x_2, \ldots, x_n)}{x_1 \partial x_2 \ldots \partial x_n} \]

The fractional polynomials (5), which can describe a partial dependence of \( n \)-input variables of each neuron, are applied as terms of the DE (6) composition. They partly create an unknown multi-parametric non-linear function, which codes relations of input variables. The numerator of (5) is a polynomial of complete \( n \)-input combination degree of a single neuron and realizes a new function \( z \) of formula (4). The denominator of (5) is a derivative part, which gives a partial mutual change of some neuron input variables and its polynomial combination degree \( m \) is less then \( n \). It arose from the partial derivation of the complete \( n \)-variable polynomial by competent variable(s).
III. IDENTIFICATION OF SIMPLE LINEAR DEPENDENCIES

Consider a very simple dependence of 2-input variables, which multiplicity is constant (e.g. \(=2\)). D-PNN will contain only 1 block of 2 polynomial neurons \(7(8)\) as terms of DE (Figure 4.). As the input variables don’t change constantly, it is necessary to add both terms (fractional polynomial of derivative variable \(x_i\) and \(x_j\) in the DE (block). D-PNN will learn this relation easily according to samples of the training data set by means of genetic and evolution algorithm (GA) [7].

\[
y_i = w_1 \left( a_0 + a_{x_1} + a_{x_1 x_2} \right)^{\frac{1}{2}} \]
\[
y_j = w_2 \left( a_0 + a_{x_1} + a_{x_1 x_2} \right)^{\frac{1}{2}} \tag{7, 8}
\]

Consider a more complicated linear dependence, where 2 variables depend on a 3rd. For example sum of the first 2 variables equals the 3rd variable \((x_1 + x_2 = x_3)\). The complete DE (for derivatives 1 and 2-combinations of block) consists of 6 terms (neurons) but only 3 of them will be enough for derivative terms \(x_1, x_2, x_3\) \((9), x_1 x_2, (10), x_1 x_3, (11)\). If other terms (neurons) are added, the D-PNN will work amiss (see Figure 5). Two-variable combination polynomials of numerators \(7(8)\) can be also applied, which could improve the D-PNN functionality and increase the number of the DE terms. This 3-variable dependence is described by more complicated exponential functions. The D-PNN as well is charged by the partial derivatives of main and inner functions \((15)\) [6].

\[
y_i = \phi_i(X) = \phi_i(x_1, x_2, ..., x_n) \quad i=1, ..., m \tag{12}
\]

\[
F(x_1, x_2, ... , x_n) = f(y_1, y_2, ... , y_m) = f(\phi(X), \phi(X), ..., \phi(X)) \tag{13}
\]

\[
\frac{\partial F}{\partial x_n} = \sum_{i=1}^{m} \frac{\partial f_i}{\partial y_i} \frac{\partial y_i}{\partial x_n} + \frac{\partial f_i}{\partial y_i} \frac{\partial y_i}{\partial x_n} + ... \frac{\partial f_i}{\partial y_i} \frac{\partial y_i}{\partial x_n} \tag{14}
\]

Multi-layered D-PNN creates compound polynomial functions. Main exponential functions of higher layers “carry” some secondary functions of previous layers, describing the partial relations of its variables. From mathematical point of view the 1st hidden layer forms the inner functions, which substitute the input variables of 2nd hidden layer neuron and block polynomials - the outer functions. Provided this assumption we are able to calculate the partial derivatives of compound functions by variables of previous layers as DE terms \((14)\), from the inner functions \((12)\) of an outer function \((13)\). These compound DE terms are formed as products of partial derivatives of main and inner functions \((15)\) [6].

\[
y_i = \phi_i(X) = \phi_i(x_1, x_2, ..., x_n) \quad i=1, ..., m \tag{12}
\]

\[
F(x_1, x_2, ... , x_n) = f(y_1, y_2, ... , y_m) = f(\phi(X), \phi(X), ..., \phi(X)) \tag{13}
\]

\[
\frac{\partial F}{\partial x_n} = \sum_{i=1}^{m} \frac{\partial f_i}{\partial y_i} \frac{\partial y_i}{\partial x_n} + \frac{\partial f_i}{\partial y_i} \frac{\partial y_i}{\partial x_n} + ... \frac{\partial f_i}{\partial y_i} \frac{\partial y_i}{\partial x_n} \tag{14}
\]

\[
\frac{\partial F}{\partial x_n} = \sum_{i=1}^{m} \frac{\partial f_i}{\partial y_i} \frac{\partial y_i}{\partial x_n} + \frac{\partial f_i}{\partial y_i} \frac{\partial y_i}{\partial x_n} + ... \frac{\partial f_i}{\partial y_i} \frac{\partial y_i}{\partial x_n} \tag{15}
\]
Each block of the D-PNN forms partial DE terms utilizing its basic and extended neurons. Single adjustable polynomial (P in Figure 6.) without derivative part creates the block output (applying in the next hidden layer) but the neurons are applied only for the total DE composition. The blocks of the 2nd and the following hidden layers create compound terms (CT) of the DE using their additional extended neurons, outputs and inputs of back connected blocks of previous layers. Consider for instance the 1st block of the last hidden layer, which takes its own neurons as 2 basic terms (16) of the DE (6). Subsequently it creates 4 extended terms of the 2nd (previous) hidden layer variables, using reverse output polynomials and inputs of 2 bound blocks. It creates 4 fractional compound terms of the DE for 4 derivative input variables of previous hidden layer using derivations of compound and inner functions (17). As couples of variables of the inner functions \( \phi_i(x1, x2) \) and \( \phi_j(x3, x4) \) differ from each other, their partial derivations are \( = 0 \) and so the sum (15) will consist only of 1 term.

\[
y_i = \frac{w_i}{\sum_j w_i} \left( \frac{x_{2j} + x_{2j} - x_{2j} - x_{2j}}{2(b_j + b_j)} \right)^k = \frac{w_i}{2(b_j + b_j)} \left( x_{2j} \right)^k
\]

Subsequently it creates 4 extended terms of the 2nd (previous) hidden layer variables, using reverse output polynomials and inputs of 2 bound blocks. It creates 4 fractional compound terms of the DE for 4 derivative input variables of previous hidden layer using derivations of compound and inner functions (17). As couples of variables of the inner functions \( \phi_i(x1, x2) \) and \( \phi_j(x3, x4) \) differ from each other, their partial derivations are \( = 0 \) and so the sum (15) will consist only of 1 term.

\[
y_i = \frac{w_i}{\sum_j w_i} \left( \frac{x_{2j} + x_{2j} - x_{2j} - x_{2j}}{2(b_j + b_j)} \right)^k = \frac{w_i}{2(b_j + b_j)} \left( x_{2j} \right)^k
\]

The previous layer block reverse outputs are used to create necessary partial derivations of the outer and inner functions (of polynomials) of differential neurons (17). Likewise compound terms can be created for the 1st hidden layer (18). The 3 linked blocks forming 8 terms of the DE were attached to the presently adjusted block. This can be performed well by a recursive algorithm. It was not every term that was used in the complete DE; some of them were necessarily left out. This indicates “0” or “1” in the neurons of blocks and is easy to use them as genes of GA. Parameters of polynomials are represented by real numbers. A chromosome is a sequence of their values, which can be easy mutated. The D-PNN’s total output \( Y \) is the sum of all active partial DE term values according to (19), which the present active amount \( a \) can be built in.

\[
Y = \frac{\sum y_i}{a} \quad k = \text{total amount of DE terms}
\]

It can be seen, that the 3-variable D-PNN (Figure 6.) substantially consists of 3 overlaying “wedge” networks (WN), each going back from the blocks of the last hidden layer and gradually attaching to the derivative variables of previous layers. The D-PNN of the 4 dependent input variables using 2-combination blocks will have totally 6 blocks of all input combination couples in the 1st hidden layer. The number of combinations for all variables increases enormously each next hidden layer. This could be solved by applying WNs, as only some of the blocks are created and used. The total amount of D-PNN’s hidden layers could equal at least to the number of input variables (i.e. 4), as it must be able to create each combination of which and to reach back all derivative variables of the 1st layer. So WNs of the 1st hidden layer will involve min. 4 random blocks, consequently in the 2nd layer will contain min. 3 blocks, etc. This way the number of all WN blocks decreases each next hidden layer until is reached just 1 block. D-PNN will have several overlaying WNs partly in the layers again. Some WN layers overlay each other and so the blocks can be used several times by different WNs (Figure 7.). The blocks of the 2nd and following hidden layers can be reconnected and this could compensate missing combination blocks. The connections of the complete 1st hidden layer blocks are fixed. Likewise the previous 3-variable D-PNN type does, it can construct the partial fractional terms of the DE from back-connected blocks of previous layers. All WN blocks attach back gradually the derivative variables of previous layers. The searching space contains a great amount of local error solutions, which GA can finish easily. This problem is caused by a lot of possible combinations of block inputs and composed DE terms (only some of them may be employed), which selection is a critical phase of the D-PNN’s construction, besides the simultaneous parameter adjustment [10].

![Fig. 7. “Wedge” networks of the 4-variable 2-combination D-PNN](image)

The D-PNN of the 6 dependent input variables using 2-combination blocks will have totally 15 blocks of all input combination couples in the 1st hidden layer and 6 hidden layers. However in experiment with right triangles (Figure 9 and Figure 10.) it could be sufficient with 4 hidden layers again, because there is the maximum of 4-variable dependence to identify.

**IV. EXPERIMENTS**

4-variable D-PNN is able to identify row/column or diagonal dependence of chess pieces (Figure 8.). Input vector is formed by their \( x, y \) positions (row, column). If the white rook checks the black bishop their \( x \) or \( y \) positions equal and this can D-PNN learn to identify. Another relation occurs if the black bishop checks the white rook, the sum or difference of their \( x \) and \( y \)-positions are equal \( A_x + A_y = B_x + B_y \) or \( A_x - A_y = B_x - B_y \). Table 1 and Table 2 show network responses to dependent and
independent variables of input vectors. Training data set can consist of following 5 data samples x and y-positions of the chess pieces (the chessboard was enlarged) [9]:

\{1+22=20+3\}, \{16+1=2+15\}, \{34+3=30+7\}, \{60+30=42+48\},
\{5+25=2+28\}, \{2+5=4+3\}

![Fig. 8. Relations of chess pieces](image)

TABLE I. RESPONSES TO RANDOM INPUT VECTORS WITH 4 DEPENDENT VARIABLES

<table>
<thead>
<tr>
<th>Input vector</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>31 + 46 = 11 + 66</td>
<td>0.9654</td>
</tr>
<tr>
<td>18 + 57 = 24 + 51</td>
<td>1.0273</td>
</tr>
<tr>
<td>46 + 30 = 25 + 51</td>
<td>1.0018</td>
</tr>
<tr>
<td>58 + 44 = 17 + 85</td>
<td>1.0006</td>
</tr>
<tr>
<td>14 + 25 = 1 + 38</td>
<td>0.9882</td>
</tr>
<tr>
<td>4 + 23 = 8 + 19</td>
<td>1.0031</td>
</tr>
</tbody>
</table>

TABLE II. RESPONSES TO RANDOM INDEPENDENT INPUT VECTORS OF 4 VARIABLES

<table>
<thead>
<tr>
<th>Input vector</th>
<th>Output</th>
<th>Input vector</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>46 + 55 = 23 + 108 (=10)</td>
<td>1.2041</td>
<td>25 + 45 = 31 + 39 (=10)</td>
<td>0.8461</td>
</tr>
<tr>
<td>11 + 55 &lt; 27 + 49</td>
<td>1.1865</td>
<td>57 + 54 &gt; 43 + 58</td>
<td>0.9564</td>
</tr>
<tr>
<td>56 + 49 &lt; 5 &lt; 110</td>
<td>1.1213</td>
<td>15 + 42 &gt; 12 + 35</td>
<td>0.9179</td>
</tr>
<tr>
<td>8 + 29 &lt; 21 + 26</td>
<td>1.1587</td>
<td>49 + 47 &gt; 52 + 34</td>
<td>0.9162</td>
</tr>
<tr>
<td>24 + 34 &lt; 2 &lt; 87</td>
<td>1.1917</td>
<td>31 + 32 &gt; 37 + 16</td>
<td>0.8704</td>
</tr>
<tr>
<td>3 + 39 &lt; 26 + 26</td>
<td>1.4803</td>
<td>45 + 56 &gt; 30 + 61</td>
<td>0.9515</td>
</tr>
<tr>
<td>35 + 9 &lt; 18 + 36</td>
<td>1.2038</td>
<td>27 + 58 &gt; 45 + 30</td>
<td>0.9226</td>
</tr>
<tr>
<td>29 + 6 &lt; 37 + 8</td>
<td>1.1653</td>
<td>44 + 34 &gt; 36 + 32</td>
<td>0.9267</td>
</tr>
<tr>
<td>60 + 43 &lt; 41 + 72</td>
<td>1.0526</td>
<td>39 + 25 &gt; 30 + 24</td>
<td>0.9174</td>
</tr>
<tr>
<td>5 + 40 &lt; 40 + 15</td>
<td>1.2457</td>
<td>59 + 39 &gt; 51 + 37</td>
<td>0.9414</td>
</tr>
</tbody>
</table>

A separating plane could be noticed, detached from the relative “classes”, which have the same characteristic (if the sum of the 1st couple is less then it should be, the output is less then the desired round and other hand round). D-PNN can be trained only with small input-output data samples (likewise the GMDH polynomial neural network does) to learn any dependence [8].

6-variable D-PNN can learn to identify (generalize) a changeable shape (e.g. triangle) regardless of its size or position in the input matrix (Figure 9.). Input vector of the D-PNN is formed by x, y (row, column) coordinates of the 3 triangle apexes A, B, C. The dependence of points A, C is diagonal, A, B vertical and B, C horizontal. As there are simultaneously occurred 3 types of point relations, it is necessary to increase the number of blocks of D-PNN’s hidden layers. Training data set can consist of following 5 data samples – right equal-based triangles:

\{6,44,11,44,11,49\}, \{22,13,32,13,32,23\}, \{10,30,25,30,25,45\},
\{3,20,23,20,23,40\}, \{25,50,50,50,75\}

Testing random right triangles must keep the apexes A, B, C dependent to be correctly recognized by D-PNN. Table 3 shows responses of trained network to dependent right triangles. Table 4. applies only vertical deformations (+ and –) of right triangles in C apexes (Fig.10, down), to be shown a transparent separating plane, detaching the relative triangles.

![Fig. 9. Dependent right equal-sided triangle shapes](image)

![Fig. 10. Independent deformed right triangles in apexes A and C](image)
Only linear dependencies of variables have been assumed for simplicity in the examples presented. If there is an occurrence of a non linear dependence of the input data, the square power exponent variables would extend combination polynomials of neurons and applied also as competent derivative terms (20). The denominators of (20)(21) result from degree partial derivations of 2(variable compound function for simplicity in the examples presented. If there is an unknown dependencies of the data variables could also be considered. This could be regarded as a pattern of abstraction, similar to that utilized by the human brain, which applies the approximation with time-delayed periodic activation functions of biological neurons in high dynamic system of behavior. D-PNN is a new type of neural network, which performs identification based on any unknown generalized relations of input variables. D-PNN forms its functional output as a composition of differential equation terms (which describe a system of dependent variables) from rational integral functions. The problem of the multi-layered D-PNN construction reside creates every partial combination term for a complete DE in utilizing some fixed low combination degrees (2, 3), while the amount of variables is as a rule higher.

## VI. CONCLUSION

Artificial neural networks in general respond to related patterns with a similar output. They identify input patterns on the bases of their relationship. Likewise, the identification of unknown dependencies of the data variables could also be considered. This could be regarded as a pattern of abstraction, similar to that utilized by the human brain, which applies the approximation with time-delayed periodic activation functions of biological neurons in high dynamic system of behavior. D-PNN is a new type of neural network, which performs identification based on any unknown generalized relations of input variables. D-PNN forms its functional output as a composition of differential equation terms (which describe a system of dependent variables) from rational integral functions. The problem of the multi-layered D-PNN construction reside creates every partial combination term for a complete DE in utilizing some fixed low combination degrees (2, 3), while the amount of variables is as a rule higher.

### REFERENCES


