Performing coordinate transformation by artificial neural network

Arpad Barsi

Künstliche neuronale Netze sind heute eine der größten Erfindungen. Sie ähneln tierischen oder menschlichen Gehirnen und werden in vielen Anwendungen eingesetzt. Auch bei der in der Geodäsie und in der GIS-Praxis üblichen Koordinatentransformation können sie verwendet werden.

Introduction

One of the most frequently used tools in surveying, geodesy and geoinformatics is the transformation of coordinates. These tools have therefore many realizations, which intend to perform the operation possibly in an optimal way. The transformation can be optimized for giving the highest accuracy, running as fast as possible or it should be a universal technique, which can be a gateway between several systems. The need to have such a tool results that we can have different methods to execute this task.

Artificial neural network technology could be one of the unnumbered ways, because these solutions have nice function approximation feature. This feature gave the idea to design and test a neural network, which can help in this kind of work.

Following this way of thinking we can imagine the transformation as a function, which means connection between two projection and coordinate systems. Our job is to find the adequate function between them. In the concrete realization the coordinate systems were the WGS84 system and the special Hungarian projection system. It must be underlined that another systems could be used as well; it was a test, where dozens of points were given for the transformation and test purposes.

WGS84 is the coordinate system of Global Positioning System (GPS). It is a geocentric orthogonal system, where points are defined by X, Y and Z values. These coordinates are mostly large numbers, people cannot notice, where the point could be found (on the Earth surface, above it or below, is it a mountain or a flat point etc.) Another system is a very special projection system. It is the official projection of Hungary, called National Unified Projection System (NUPS). The essence of this system is an IUGG 1967 reference ellipsoid, which is oriented for best fitting to the area of Hungary. The projection will continue with the application of a second order fitting Gaussian sphere; all points are projected from the ellipsoid to this sphere. After the sphere, a reduced, oblique cylinder is used to map the projected points. This "doubled projection" has homologous and small projection errors in the whole country. This special system is very suitable to demonstrate how the coordinate transformation works with fully different systems.

Theoretical background

Artificial neural networks have

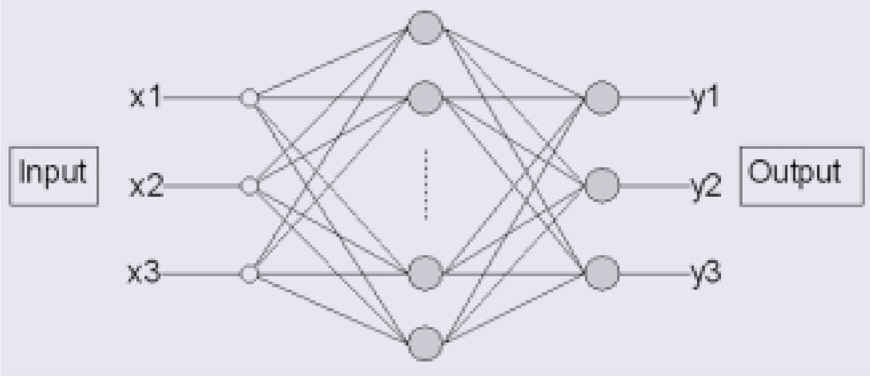

Fig. 1: The structure of the two layer neural network

many types. Maybe the most used category is the feed-forward networks [1] [3]. Several network types were tested in the transformation, now **I** limit the presentation just for the successful one. This was a radial basis neural network.

Radial basis neural networks have two layers, where the processing elements can be found (gray dots in Fig. 1). The processing elements are the neurons in the analogy to the human and animal brain. The neuron (similarly as in the nature) has a simple and uniform algorithm to calculate the output. The neurons in the first layer will produce the output after the next formula:

$$\mathbf{O_1} = f(\|\mathbf{W_1} - \mathbf{I}\| \cdot \mathbf{b_1})$$

where **I** is the input vector, $\mathbf{W_1}$ its weight matrix, $\mathbf{b_1}$ its bias vector, the so-called transfer function. The multiplication in the above term (·) is an element-by-element multiplication, not as in matrix algebra. The radial basis neural network takes the following formula for transfer function:

$$f(x) = \exp(-x^2)$$

It can be noticed that the transfer function is a well-known Gaussian bell curve. In the calculation the network practically compares the input values to its weights, calculates a distance, which is multiplied by a bias vector. The output ($\mathbf{O_1}$) is therefore a vector, which is the transfer function applied for.

The second layer gets the output vector of the first one, and will calculate the network's output (**O**) as following:

$$\mathbf{O} = \mathbf{W_2} \cdot \mathbf{O_1} + \mathbf{b_2}$$

where $\mathbf{O_1}$ is the input for the calculation (taken from the first layer), $\mathbf{W_2}$ the weight matrix of the second layer and $\mathbf{b_2}$ the bias vector of the second neuron layer. In this case, several authors [2] mention a transfer function, which is a simple linear function, giving the same output as the input was.

The parameters of the radial basis neural network are the weight matrices and bias vectors of both layers. Their values are defined during the training phase.

The training requires data points,

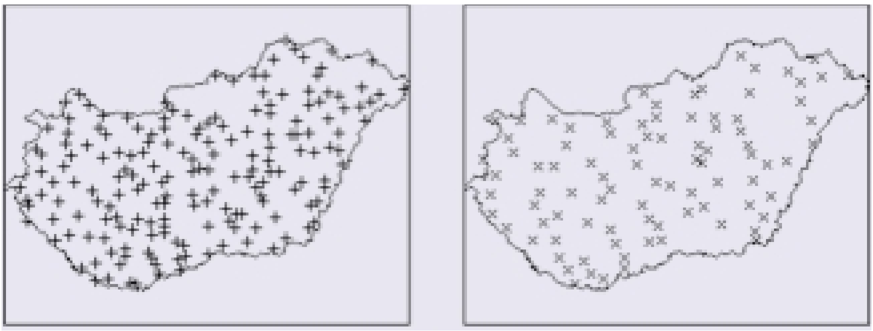

Fig. 2: The distribution of the training (+) and test (x) points in Hungary

where the coordinates of both systems are known.

Mathematically the training is an optimization problem, where the network parameters are to be solved, while the error of the network output must be minimal. To measure the network error (Δ), the next formula is introduced:

$$\Delta = \frac{1}{n} \sum_{i=1}^{n} \delta_i^2$$

where n is the number of network's output, δ the difference between the required and the calculated network output. This mean squared error is very fast to evaluate during the training.

The training is an iterative procedure, where the process starts with a single neuron on the first layer. The error is calculated and then the network parameters are defined propagating back the effect of error to every parameter. This is why the differentiation is so important! If the network error is too high, the steps are to be repeated. The iteration will add neurons to the layer, until the desired error value is reached, or the maximum number of neurons are added. The maximum amount of processing elements is the number of data points. (In this case the network learns the data set perfectly, but loose the ability to work with other points.) The method ensures the minimum number of required neurons. This is advantageous to get a suitable network.

After the training the network parameters are set, and the new coordinate transformator is ready for work.

Realization

Before the above described training was executed, the coordinates of both systems were taken and are transferred into the system of the center of gravity. This step makes easier to handle the coordinates and the condition of the training is better. (This transfer has "tradition" in the history of coordinate transformation, the mention of Helmert's method is enough.)

The training set contained 180 fix points, where GPS measurements were executed and their NUPS-coordinates were taken from earlier geodetic measuring campaigns. There were additional 90 points, which didn't take part in the training, have known NUPS coordinates and GPS measurements were there. This point set was for testing the resulting neural network (Fig. 2).

The training was executed, some regulating parameters (concerning the iteration and training) were set and the network was ready after some cycles. The resulting network had 22 neurons on the first layer. The second layer must be fixed, because the artificial neural network must produce 3 coordinates (y, x, h – following the traditional Hungarian nomenclature), so this layer had 3 neurons.

Table 1: The maximum and average differences in the coordinates of the test points

	Maximum difference [m]	Average difference [m]
X	0.176	0.050
Y	0.192	0.061
Z	0.369	0.129

After the test points the transformation network could be tested. During the test the differences in all coordinates were computed, then the length of the difference vectors, too. The highest and mean differences are shown in Table 1.

The maximal difference vector length was 0.394 m. The mean square error of the network was 10^{-10}.

As it can be seen from the above table, the average and maximum errors are not in the same category: the average is about 3 times better, than the maximum error. Further interesting notice is, that the horizontal coordinates are much better than the vertical one. It is supposed, that this is caused by the different horizontal and vertical measurements and adjustment of the traditional network. At this point further analyses should be done together with geodesy colleagues.

The difference vectors were nice base to create an accuracy map, which can demonstrate the network's accuracy. For this reason Fig. 3 was plotted.

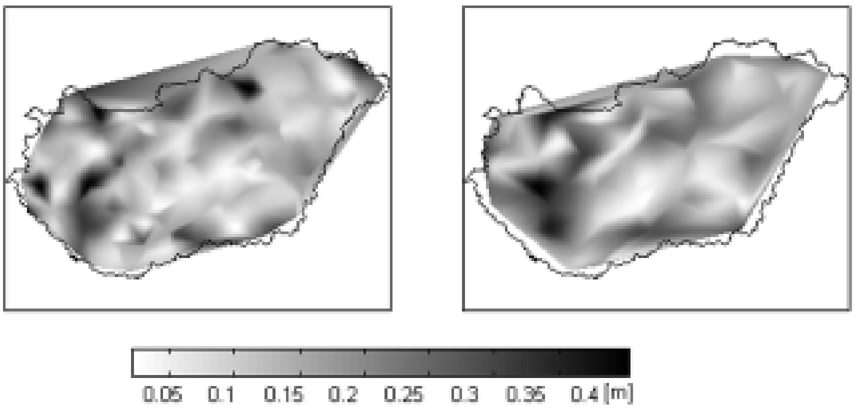

Fig. 3: Interpolated distribution map of the network error in the whole country

Analyzing the map, it could be noticed, that the error distribution has the similar picture as Hungary's topography. The correlation between the height and the sensible error was checked by executing the training and test with other coordinate orders. The result was the same: the order of the coordinates doesn't play a role in the transformation. The only reason therefore is a strong correlation with the height measurement and adjustment.

The accuracy analysis pointed that the transformation network is suitable for topographic and similar medium to small-scale works.

In order to make usable the developed transformation method, software – called Neutra – was written with graphical user interface (GUI). Applying Neutra, the user can choose between the Single Point Transformation and the Multi Point Transformation mode. Single Point Transformation makes possible a distinct point with WGS84 coordinates to be converted into NUPS coordinates. Fig. 4 shows this user interface. (EOV is the Hungarian abbreviation of NUPS.) Multi Point Transformation reads a standard ASCII-file containing the WGS84 coordinates and produces a similarly structured output file. With this option, the software can be linked to other geodetic or GIS applications.

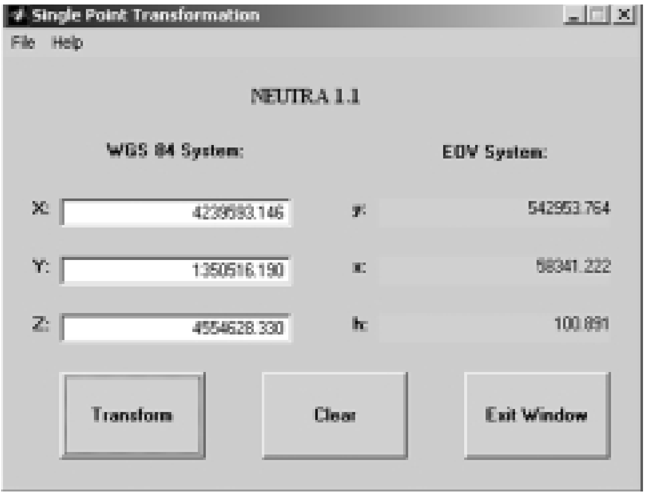

Fig. 4: The graphical user interface of the Single Point Transformation mode in Neutra

Conclusion

Artificial neural networks as function approximators have the possibility to be an adequate tool in coordinate transformation. To prove this hypothesis several type of neural networks were designed. One successful version was the radial basis neural network type.

The training of the network was executed by points having known coordinates in both input and output systems. After the training phase the resulting network was tested, which gave that this new transformation method is suitable for practical works.

I want to underline that the above presented artificial neural network is just a way of solution, not only a special transformation. One of the best features of using neural networks is that they contain processing elements with the same functionality. The neurons must evaluate a function in every case. As it was mentioned yet, the only parameters of the network are the weight and bias values. If someone could determine a parameter set for other two projections or coordinate systems, the network will work. This means that artificial neural networks realize a so-called „drawer discipline" i.e. the currently required parameter set could be taken from the memory. The network structure doesn't need to be altered; just the parameter set must be changed!

Because of this fact, we can state, that the method of artificial neural network is a universal computing tool, in contrast to the special projection calculation, which are built of specific projection equations.

Neural networks have a nice feature that is they are in vicinity with hardware development. Neural networks can be realized with low efforts on hardware platforms, in form of computer and measuring instruments cards, panels. The network can have "hard parameter set", i.e. it is stored and fixed on the card. Another way is to have a "soft parameter set", which means changeable parameters. The advantage of this later one is the enhanced flexibility and the ability to update.

At the end two interesting and possible applications must be mentioned. The first is the usage with mobile phones, which contain a GPS receiver, calculates the position of the phone and can be used in emergency situations or in searching different services (e.g. bank, restaurant or drug store). Furthermore the system could be employed in car navigation and fleet management. Vehicles having GPS stations can report their position to the dispatcher center, where the local GIS and road information (motels, borders, petrol stations etc. and slippery road segments, dangerous curves, traffic or accidents etc.) can be queried and send back for the driver. The whole information amount is sometimes managed by local authorities, which can use their own mapping system. This case could be very helpful for example in the international cargo.

Acknowledgement

The author is thankful for the kind support of the Bólyai Scholarship of the Hungarian Academy of Sciences.

Reference

[1] BRAUSE, R. (1995): Neuronale Netze, B. G. Teubner, Stuttgart

[2] DEMUTH, H. – BEALE, M. (1998): Neural Network Toolbox – For Use with MATLAB, User's Guide, The MathWorks Inc., Natick MA

[3] ROJAS, R. (1993): Theorie der neuronalen Netze – Eine systematische Einführung, Springer Verlag, Berlin

Adress of the author:
Dr. ARPAD BARSI, PhD, Ass. prof.
Budapest University of Technology and Economics, Department of Photogrammetry and Geoinformatics
H-1111 Budapest, Muegyetem rkp.3., Hungary
barsi@eik.bme.hu

Zusammenfassung

Transformieren von Punkten in verschiedenen Koordinatensystemen ist eine alltägliche Arbeit. Für die Lösung dieses Problems sind verschiedene traditionelle Methoden angewendet. Wegen der Funktionapproximationsaspekte der künstlichen neuronalen Netzwerke eignen sie sich auch für diese Aufgabe. Die Forschung hat gezeigt, dass neuronale Netzwerke in Verbindung von zwei völlig anderen Projektionen und Koordinatensystemen erreichbar sind. Das Netzwerk hat genügend Genauigkeit für topographische und kleinmaßstäbliche Zwecke.

Summary

Transformation of points in different coordinate systems is an everyday task. To solve this problem, different traditional methods are applied. Because of the function approximation feature of artificial neural networks, they are also able to execute this job. The research has proved, that a neural network can be designed and trained for connecting two totally different projections and coordinate systems. The network has enough accuracy for topographic and small-scale purposes.

