České vysoké učení technické v Praze
Fakulta biomedicínského inženýrství

Czech Technical University in Prague
Faculty of Biomedical Engineering

RNDr. Ing. Marcel Jiřina, Ph.D.

Využití neuronových sítí při zpracování biomedicínských dat

Utilizing of Neural Networks by Processing of Biomedical Data

Summary

The biological neural system of the living organism presents centre of attention of a number of branches of knowledge. We do not have to study the neural system or its function only, we also can be inspired by this system and try to create more or less exact prototype with similar properties as the real specimen has. These models can do well in a number of technical branches particularly, but also in humanitarian ones by solving a number of problems, as e.g. knowledge extraction from data, prediction of time series, creation of implicit models of the examinated systems etc. The other direction, where the science of the neural systems of living organism proceeds is an exact description, i.e. simulating and/or modeling of real biological neurons including description of their interconnection to form the neural networks.

The theme of the lecture is artificial neural networks and their relation to the biomedical Engineering. This lecture aims especially at the correct approach how to utilize the artificial neural networks, i.e. how to preprocess the data and how to evaluate the achieved results, further at the exemplary utilizing of the neural networks in biomedical engineering. There are discussed the possibilities to utilize neural networks by processing of single-dimensional signals and two-dimensional images. To the classical tools of the present-day belong the expert systems. The neural networks can be successfully utilized as a suitable tool that enables to extract coherences in data and to represent the consequent rules.

Souhrn

Biologická nervová soustava živých organizmů je ve středu zájmu celé řady vědních oborů. Nervovou soustavu resp. její funkce nemusíme jen zkoumat, ale můžeme se jí inspirovat a vytvářet více či méně přesné modely, od kterých si slibujeme, že budou mít podobné vlastnosti jako jejich reálné vzory. Takovéto modely pak slouží v řadě zejména technických, ale i humanitních oborů při řešení řady problémů, jako je např. extrakce znalostí z dat, predikce časových řad, vytváření implicitních modelů zkoumaných soustav apod. Jiným směrem, kterým se ubírá věda o nervových systémech živých organizmů, je exaktní popis a tedy modelování reálných biologických neuronů včetně popisu způsobu jejich propojení do nervových sítí.

Tématem přednášky jsou umělé neuronové sítě a jejich vztah k biomedicínskému inženýrství. Přednáška se zaměřuje zejména na přístupy jak správně využít umělé neuronové sítě, tj. jak předzpracovat data a vyhodnotit výsledky a dále na ukázkové příklady využití neuronových sítí v biomedicínských aplikacích. Jsou diskutovány možnosti využití neuronových sítí při zpracování jednorozměrných signálů a dvourozměrných snímků. Ke klasickým nástrojům současné medicíny patří expertní systémy. Neuronové sítě lze s úspěchem využít právě jako nástroj, který umožňuje extrahovat závislosti v datech a reprezentovat pravidla.

Klíčová slova: neuronové sítě, biomedicínské inženýrství, urychlení učení, předzpracování dat, vyhodnocování výsledků, modelování

Keywords: neural networks, biomedical engineering, training speedup, data preprocessing, result evaluation, modeling

Table of Context

1. Artificial Neural Networks

We can encounter the interpretation Neural network at first in connection with biological nervous systems of living organism and secondly as with an artificial system created and framed by man. As we will be now dealing with the artificial neural networks, we should specify this conception. There is no exact definition of an artificial neural network. In spite of this, let's present following definition to delimit this phenomenon in a sufficient scope for our use.

The *artificial neural network* is a distributed (to be understood as parallel) computational system (program) consisting of sub-systems (neurons), that is inspired by neuro-physiology piece of knowledge concerning the structure and activity of neurons and neural systems of living organism. This artificial neural network is in a larger or smaller scale modeling this procedure.

There exists a number of so called paradigms that specify various types of neural networks. You can find their detailed description in the habilitation thesis. Here are also specified some author's methods that can expand the possibilities of the present neural networks. First of all it concerns the methods of first initialization of weight in Kohonen's network that enables an obvious acceleration of subject acquisition.

From the point of view of relation (of artificial) neural networks towards the biomedical engineering we can split this view into two parts: The first one is the utilization of neural networks as a mathematical models, where the properties and capabilities of those enables to solve a number of tasks we can meet on the field of biomedical engineering. As an example we can take the data evaluation, signal, images, status and/or condition – prediction etc. In this case the biological inspiration is independent of the matter to be solved. The second example is the endeavor to depict the biological phenomenon mathematically and to create in such a way more or less exact models of behavior. Such models can be used for various analyses and behavior – predictions.

In the next part of the lecture we will first aim on the immediate utilization of the neural network as a means to process the biomedical data. There is very important in this respect to aim the attention to the correct data preparation for the neural network and also to the correct evaluation of the achieved results.

1.1. The Work with Neural Networks

Especially in the field of neural networks and its utilization in the biomedical engineering there is important to consider carefully the individual steps when preparing and evaluating the data, because the achieved results can have a crucial influence on the investigated problem. We have to realize that the neural networks itself are not something like panacea, but a supporting means where especially in the biomedical sphere there is essential to understand the individual steps when processing the data.

One of the main steps in manipulation with the neural networks is the principal proposal how to process the task, i.e. to propose the individual steps to be performed. Together with this step there is connected in our case the choice of the suitable neural network, its parameter, the choice of the training algorithm for the data preprocessing and the methodology of the data evaluation.

Generally, we request the accessible data, i.e. the data acquired by experiments (measured on tangible objects) or by simulation (from models) to be as much representative as possible (as per the quality point of view) (Flexer 1994). Moreover we can request that the relative representation of individual dependences in data would reflex the reality. This request is not always accomplishable and on the other hand, we do not need and request it necessarily. The acquisition of the representative data should not be, in optimal, intruded by noise (disturbances).

As concerns the extent of data, i.e. the requirement to have them as many as possible, it is necessary to have as many as possible of patterns (i.e. acquired marked vectors). From the practical and extensional point of view are welcome ranges of ten thousands patterns. The thousands of patterns are still processible, but we can say the numbers under one thousand of patterns may create already a problem. This concerns especially the field of medicine, where each the pattern represents frequently one patient and we can acquire here an extensive set of patterns with greatest difficulty only. When considering that we need the sampling and division into the individual sets, then the situation is even much worse.

The purpose of data preprocessing is to clean the data from the undesirable influence. The main activity during the data preprocessing is to remove the noise (disturbances), data filtering, different transformations (e.g. geometrical or time transformations((Dowla, Rogers, 1995; Swingler, 1996), to reduce the dimension of entry space, i.e. neglection of not important entries or transformation of patterns into space of lower dimension (Jiřina, Jiřina, 200ř (a,b), rectification of data division, i.e. transformation from the given division to a normal division, detection and removal of wrong patterns (i.e. of those, where some marks of features fails) made by their omission from the data set or by rectification by using the mean of the nearest neighbor etc. (Masters, 1994; Stein 1993).

From the primary data set (population) there is made a selection and from here further a division into training, validation and testing (and some other) sets (a two – stage selection).

1.1.1. Selection (Resampling) of data

Nowadays there is being favored the system of three sets – i.e. the training one, the validation one and testing one (Michie, 1994). The individual patterns are being chosen into individual sets by random from the data set. In practice there is being used to divide the data into training, validating and testing set the ratio of 2:1:1. There exists also another recommendation to divide the data in the ratio of 2:1:2 (Michie, 1994).

The *training set* used to be selected as the greatest one, because we consider it as the most important one to practice the neural network thus we have here enough of pattern to practice in order to secure a quality training. To prevent *overfitting*, i.e. to prevent the neural network to adopt its parameters to thoroughly and consequently to show a bad behavior on the testing data, we use so called *validation set*, that is chosen as slightly smaller compared to the training one and its task is to prevent the above mentioned overfitting. As soon as the neural network is trained up, the testing set is being used to verify the quality of the learned network.

Another sophisticated method is e.g. *k-fold cross validation, leave-on-out* (Hecht-Nielsen, 1990; Rizzo, Dougherty, 1994; Lavesson, 2003), *double cross-validation*, (Mosteller, Turkey, 1997), *bootstrapping* (Efron, Tibshirani, 1993; Breiman, 1996; Lavesson, 2003; Flexer, 1994).

1.1.2. Normalization of data

The non-normalized data can cause serious problems during the network training. E.g. in case of multi-layer perceptron network there exists a danger of premature saturation of latent neurons (the output data will acquire values close to zero or vice versa values close to one, when using the sigmoid) what slows down or event stops the learning.

There exists a number of very demanding normalization methods, what the computation concerns (Masters, 1994; Swingler, 1996; Dowla, Rogers, 1995). Most frequently there is being used a linear normalization to interval $< 0; 1 >$, or pertinently normalization to a zero – mean and unitary dispersion variance. There exists also more sophisticated non-linear method, which e.g. try to change the scale in the data in such a way so that the more important extent included in the data is highlighted. As an example we can take the normalization of a time progressions or of a small set of data.

1.1.3. Methods of choice of features

The methods of feature's choice serves, similar as the sensitivity analyses do, for selection of suitable entries (features, attributes) for the neural network from the set of all possible accessible entries. These methods are based mostly on the principle of selection of different entry combinations and on the evaluation of their convenience.

There is e.g. being used the method of agglomeration analyses, where is being looked for the combination of feature groups creating separated agglomerations so that we can during the classification easily differentiate the individual classes. We can also successfully use here the genetic algorithm and generally evolutionary practice that optimizes the choice of groups with appropriate features. A classical method of the feature choice are so called *forward selection, forward stepwise selection*, or *backward elimination, backward stepwise selection*, when we step by step and temporarily eliminate the influence of individual features and we follow the influence of the other features on the learning quality of the neural network. Another optional possibility is to analyze the properties of data spreading in the space and based on that to arrange the features according to their importance and to select the most suitable ones that bring the most of new information to the already existing feature's choice.

1.1.4. Missing data

When performing the measurement on a real system or in case we compile the real measured data it might come up that some of the data, we measure, are not at disposal. An example: we forget to read in the right moment of time the data or there might be a blackout or damage of the measuring device. These data are to be anyhow processed, especially to unify their appearance. In case of missing data (here come also data where we know they were wrongly measured) we have to decide what to do with the patterns with incomplete information.

In case there are many measured data and there is just a few of the damaged (incomplete) data, we can overlook them. An implicit precondition is that those damaged patterns do not depend on the data itself. In case there is only relatively few of the measured patterns and each the data is very valuable there might be a solution to replace the missing link (data) by a mean value of the given variable from the other not damaged patterns. This approach might lead, of the other hand, to a problem. A better solution is to approximate the given data and the to deduct the missing link (data) from the appropriate regression function.

In case we speak here about discreet values, that have a meaning of a class identifier or in case these are the nominal values themselves, it is possible to replenish the values on the base of frequency of those data in other patterns or by random generation of the given value from the appropriate discreet categorization.

1.2. Evaluation of Classification Quality

In a number of applications a certain inaccuracy in the result's interpretation does not mind so much, but in field of biomedical engineering it becomes extraordinarily important as the correct evaluation of the results might be decisive for the human health or even life.

There exists a number of different approaches and methods to evaluate the correctness of evaluation. To the most widespread and thus generally understandable method belongs so called ROC – analyses, e.g. (Metz, 1978). Further, we'll also deal with statistical methods (e.g. cross-validation, quantitative characteristics) function evaluation (qualitative functions) and with entropy.

1.2.1. ROC - Analysis

The abbreviation ROC stands for **R**eceiver **O**perator **C**haracteristics. The ROC analysis is suitable and usable in case, when classifying into two (2) classes. We can distinguish here four (4) situations. For the patterns of 1st class (in the english bibliography known also as „positive“), that were classified correctly, i.e. they are on the right side of the discriminating boundary, we use the name „true positive“ – i.e. „TP“. For the pattern of the 1st class, that were classified wrongly into the 2nd class, we use the name „false negative“ – i.e. „FN“. Accordingly, from the point of view of the 2nd class pattern, („negative“), we can distinguish here the concept „true negative“ and „false positive“.

The ROC – curve is the relation between normalized frequencies (TP and FP) or the probabilities (P(TP) and (P(FP) of such classification, i.e. among correct classified pattern of the 1st class and wrongly classified pattern of the 2nd class. On the vertical („Y“) axis we plot on the TP – values or (P(TP). This value is also being called as *sensitivity*. On the horizontal axis („X“) we plot on the FP – values or (P(FP). The remaining adjunct of this value till one (1) bears the name *specificity* - see also the Fig. 1. The ROC curve gives us the overall characteristic of classifier. Generally, we can say: the more the curve simulates the left vertical and upper horizontal axis, or the larger is the area under the curve, the better is also the classifier.

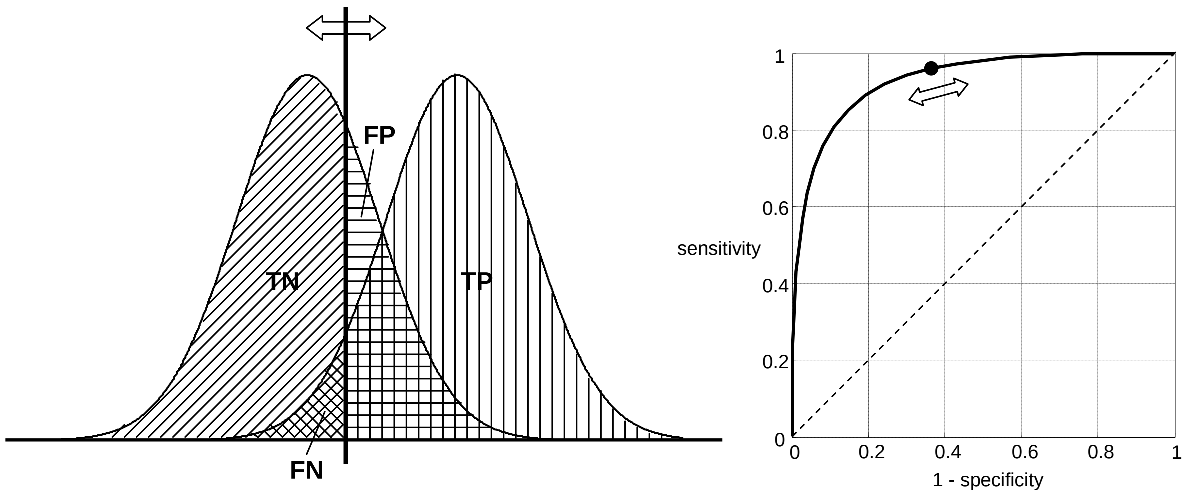


Fig. 1 ROC curve and its relation to the probability's density

1.2.2. Evaluating function for classification's evaluation

We can evaluate the quality of classification also by means of a general evaluation function that takes respect to the facts we are interested in. We call this function as a *measure function*, see e.g. (Andersson et al.., 1999). In contrary to the other already mentioned approach, this function can in its universality evaluate not only the quality of the classification from the point of view of acquired results on testing subset-fit, but it takes respect also to the fact, that similar patterns should have been classified in similarity and the classifier should be as simple as possible. This is thus a qualitative analysis. In (Andersson et al., 1999) there is proposed an evaluation function having a following form:

$$a \cdot accuracy + b \cdot similarity + c \cdot simplicity\ ,$$

where a, b and c are weight coefficient, that represents the relative representation of individual partial evaluations and the accuracy, similarity and simplicity are functions, that evaluate above mentioned requirements. According to their choice the result may get a number of different evaluating functions.

1.2.3. Statistical approach to the classification's evaluation

The methods of statistical test are a classical tool to evaluate the results from different types of analyses. In principle, they calculate different characteristic as concerns division of acquired outputs. There exist some basic statistical tests to evaluate the quality of the classification with respect to the small set of data (Dietterich, 1997). Just the small extent of the accessible data use to be very frequent in the every-day events. An example may be the scope of medical science, where is quite difficult -sometimes maybe even impossible to acquire data from thousands of people in order to be able to use the classical statistical methods.

To the basic tests, that satisfy the above mentioned tasks as concerns the small number of patterns belongs following tests: McNemara – test, test of the failure difference of the classificator, repeated selective pair t – test and k - multiple t – pair test with cross validation.

1.2.4. Evaluation of classification by means of entropy

The quality of the classification, better said of the classificator can be also evaluated from the point of view of theory of information (Kononenko, Bratko, 1991). The theory of information introduces the term of ´entropy´ as normalized measure of information included in the report (Kotek et al., 1990). The entropy is being used to evaluate and to judge of the measure of the information (i.e. of something new, fecund, what is just being conveyed) included in the evaluated message.

Just considering the frequency (probability) with which are represented the individual patterns of the given class in the respective set brings the further mentioned criterion based on the entropy. Another notable advantage of this approach is the fact, that we can not only evaluate the quality of the different classificators including the expert's approach for the same range of the problem, but we can also easily compare of different classificators for different ranges of problems. The classification's evaluation based on the Information's theory is advantageous as we consider here the probability of a representation of classes in data. And what's more – we can work even with incomplete classification. Also the interpretation of the results is natural.

2. Application Of The Neural Networks In Biomedical Engineering

We mentioned that there is necessary to take heed of data preprocessing and also of result evaluation. This issue is extra important when realizing that those result and data acquired by means of neural networks represents the base e.g. for deciding of another doctor. An insufficient, bad preparation or misunderstanding the entry data and unsound interpretation of acquired results as well may lead up to fatal conclusions and situations.

Following are common recommendations where is necessary to take heed by processing of biomedical data in neural networks – see also e.g. (Flexer, 1994; Salzberg, 1997):

- there is necessary to ensure that the training and testing sets are compiled of different patterns and thus guarantee their independency for correct evaluation of classificators; a 10-fold cross validation is recommended
- to use the validation set to prevent overfitting
- it is recommended to train more classificators on random chosen data and to consider the mean of acquired values (results)
- to enter particularly the mean, dispersion variance, and confidential intervals of acquired results
- to perform a trustworthy and complete evaluation according the recommendation mentioned in the foregoing chapter, especially the ROC analyses and also pertinently the statistical tests (binomial test, McNemara test)
- the verify the capabilities of the proposed or used classificator on the real, representative data
- to compare critically the acquired results with another classificators (with the best ones).

The neural networks are the possible and sought-after alternative to a number of methods of information processing. The connection of neural networks with other applications in the medical science in on the rapid increase, see e.g. the work concerning the citation from the field of this interconnection (Ohno-Machado, Musen, 1994; Ohno-Machado et al., 1995;

Dybowski, 2000). The neural networks have a number of properties that indicates to be very fecund. First of all it is the capability to gain knowledge from the examples and to work even with the incomplete information. The capability to generalize the acquired knowledge is welcome in the cases when we need to know the correct answer based on data, for which the network was not assigned (i.e. the capability of generalization).

In (Partridge et al., 1996) is stated a list of advantages and possibilities of neural networks in the relation to applications in the medical science. We quote here this free adopted list:

- implementation of knowledge is based on data and not on rigidly defined rules, that might be false
- interference and new cases are processed automatically owing to the capability to generalize the acquired knowledge
- the neural networks are able to predict the future values based on the foregoing data and detected trends
- analyses and diagnosis can be under way in the real, present time
- it enables a quick identification and classification of entry data
- it eliminates failures conveyed by tiredness and habits of man.

Advantages and disadvantages of neural networks are discussed almost in every treatise dealing with neural works. In (Passold et al., 1996) are summarized the advantages of neural networks in relation to medical science (again, freely adopted):

- the capability to compile a huge quantity of data
- to simulate the diffusiveness of conclusions when deducing the medical diagnosis
- to show better properties and capabilities compared to the classical statistical approach
- the capability to adopt (to learn) and adaptation to a certain problem
- the knowledge base can be easily updated.

As long as we should be actual, the neural networks are being used e.g. as a supporting item to specify the repercussion (feedback) of influence of medicaments during the medical treatment, the prediction of interaction between germs and antibiotics during treatment of infectious disease, prediction of cardio-vascular diseases, particularly heart attack, detection of diseases (e.g. arrhythmia), based on the shape change of the EKG – signal, segmentation of 2D (edge- or area-shaped) and 3D – (volume) of medical pictures, pre-processing and transformation of pictures for interpretation, detection of tumorous diseases, planning of medication etc.

2.1. Neural Expert's system

The expert's systems and knowledge systems are utilized in the medical sphere, especially as a diagnostic means (Micheli-Tzanakou, 2002). Both, the expert's and fuzzy systems do not differ in principle too much, it is more or less the point of view than a factual difference. Both the systems are based on the rules that may be compiled on a base of an analysis of compiled data. The neural networks find here a good use, as they enable to extract the dependence in data and transfer them into the form of rules.

The neural networks can be used in expert's systems in various ways. In case we will use the neural network for representation of inferential source and knowledge base, what is one of the most natural ways of neural network use as describes in the further explication, we speak in this case about (“*pure*“) *neural expert's systems*. Inferential machine and base of knowledge, performing the core of the expert's systems are in the classical expert's systems singularized. In contrary to this, in neural expert's systems represents the neural network both the inferential machine and the base of knowledge as well. In case we use the neural network in cooperation with procedures in classical expert's systems, then we speak about hybrid, (composite) neural expert's systems.

The main disadvantage of classical expert's systems is their substance. I.e. to set up the rules based on the uncertain, incomplete and only with difficulty expressible information of an expert. The neural networks can be used in the expert's systems for an automated creation of the knowledge footing based on the paradigm examples. In case the neural networks do represent direct the knowledge base, we speak about so called *neural expert's systems*. The advantage of this approach is the fact, that the knowledge footing is being created automatically by the extraction of rules based on learning from the paradigm cases. Due to the generalizing capabilities of neural network, such a neural expert's system can deduce correct conclusions even from not learned examples. The architecture of the neural expert's system is identical with the architecture of a classical expert's systems with the only difference, that the inference machine, knowledge footing (base) and facts (implicit internal knowledge) are represented by the neural network.

The knowledge in the neural expert's systems is implicit, i.e. the network gains this knowledge itself from the training paradigm and codes them in its own weight way. This is why it might by difficult to present to the user the way, how the network came to the certain conclusion. Moreover, all the derivation, the network came to, are given by capability of the network to gain these rules from the training example. This is why, that not all the rules are fully integrated and thus it might be the debased the reliance in the correct function of the neural expert's system.

The problem of the result's interpretation for the practical use by the doctors is being discussed e.g. in (Jakob, 2003). The idea of result's interpretation is being solved in such a way, that the result, that is written down as a logical statement, is simplified in a way, that the coherence are obvious to the doctor and he would be able to transfer the result into his own semblance and identify himself with the results of this expert's system. This simplification means on one side accurateness’ impairment, but it is by far more credible for a doctor, because he understands it.

2.2. Examples of neural network's application for biomedical data processing

In this part of the lecture we will demonstrate the use of some basic types of neural networks for solving of simple problems that concerns processing of biomedical signals and pictures. We will touch up here the multilayer perceptron network, related variants of Hopfield's neural network, Kohonen's maps and RBF network.

2.2.1. Prediction of signal from the seismocardiograph

A seismo-cardiograph is a device that analyses the activity of the cardio-vascular system. It detects the force that is a resultant of a number of mechanical forces, caused by the living organism. First of all are they the responses caused by the heart activity (cardio-vascular system), further by the breathing, by mechanical movements of the patient etc. (Jiřina et al., 2005 (c)).

As the measured signal itself is quite weak the other influences are very intrusive. Especially the movement of the patient causes abrupt changes of the signal's level. Likewise the breathing brings also disruptive factors into the signal. The possible way how to filter off those false data is to detect these changes and to exclude the respective part of the signal in the subsequent processing. As soon as we train the neural network in such way, that it is able to predict a normal passage, we can afterwards ask the neural network for the size of the predicted value and if that differs too much from the real signal's value (or more values in a row) it means there an adverse situation occurs.

By the preprocessing of the signal we remove first the impulse noise caused by the measuring device by means of median filter via three values. Further on we tune this signal very fine by means of averaging filter via five values. As there is in this signal still an influence of breathing and other factors, we trace this global trend and by deducting it from the filtered data we will get a signal, that oscillates around the horizontal time axis and its mean value has approximately zero – value. The signal we gained shall be further normalized to the zero mean value and unitary dispersion. The final preprocessed signal is shown in the Fig. 2.

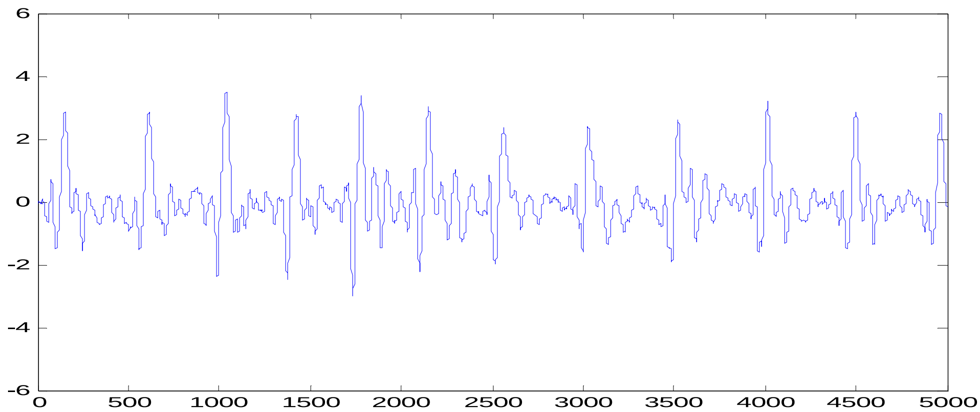


Fig. 2. Preprocessed signal from the seismocardiograph.

For the signal prediction as such we used a multilayer perceptron network with two layers of neurons. The activation functions of the neuron's first layer are sigmoids. In the second (output) layer there are used the linear activation function. The number of neurons in the first layer was 20 and in the second one 10. As paradigm for the neural network we use the data from own predicted signal (the good parts of signal only) and we arrange them into pattern. Each the pattern is being created by n - fold of the value's row, whereas for each next sample we structure the values always by one pattern shifted to the right (in the direction of the time axis). As a output (predicted signal value) we take the following sample, i.e. sample $n+1$. We have chosen the size of the sample $n = 50$. The part of the original signal together with the final predicted development is shown in the Fig. 3. We do the prediction by one step ahead.

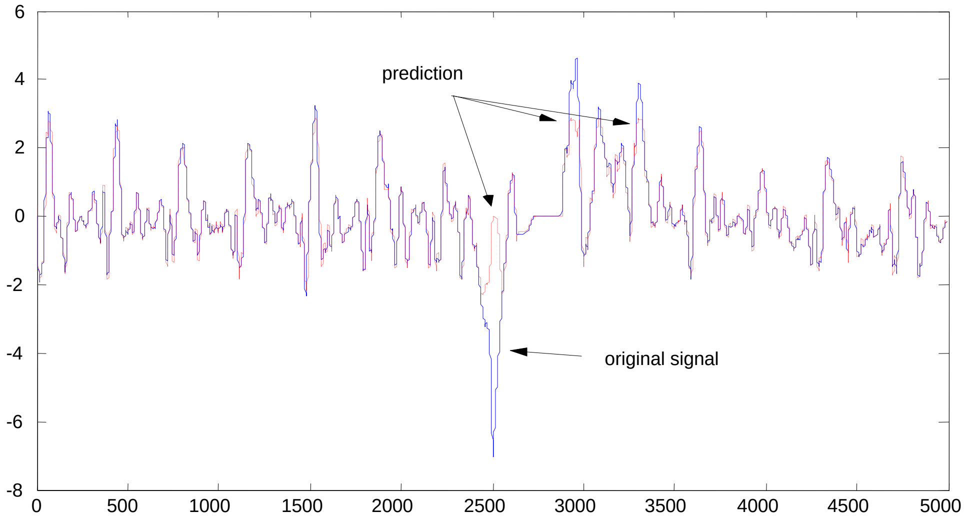


Fig. 3. A part of the original and final predicted signal form the seismocardiograph.

The use of the neural networks is in this case very advantageous, because the original signal is the repetitive one and thus not fully periodical and it is quite difficult set up the model in a classical way. Also the course in each the interval of the repetition is not always identical and there exists some small differences. The neural work enables implicitly to create a model that shapes these arduously detectable dependences.

2.2.2. Utilization of the Hopfield's network for preprocessing of medical pictures

In this part of the lecture we will aim our effort on pre-processing of the picture, specifically detection of the edges in the picture and we will show possible use of continuous version of Hopfield's network, which will find a use as an optimization method (Hopfield, Tank, 1985; Chao, Dhawan, 1994). The methods to pre-process the picture serve especially to emphasize the important information in the picture pertinently the transformation into the standard appearance for other, more sophisticated treatment. For processing of medical pictures it serves e.g. for detection of the tumor contour, visual partition of the boundary of individual organs, to follow the shape of vanes for edges segmentation etc.

The detection of edges based on the use of continuous Hopfield's network comes from the assumption that we can consider the picture as a dynamic system, that develops in time and that can be described failure (energetic) function (in case of Hopfield's network, I speak rather about the energetic function). The main idea of the Hopfield's network use is based on the idea to create a network as many neurons as many pixels are in the processed picture and to each the pixel is assigned one neuron. As we know, each the neuron is connected via the weight with all the other neurons, except itself. It means that each the pixel is in this concept interconnected with all the other pixels in the picture. The weight of the connection, i.e. the strength (importance) of the binding among the neurons (pixels) is described by the function that reflects first the mutual contrast and also the geometrical distance of individual pixels.

The initial status of each neuron corresponds to the radiance of pixels in the picture. This status alters during the processing phase and thus also the level of brightness in such a way that the energetic function is being minimized. If we in the energy function consider the mentioned geometrical distance of pixels and mutual values of pixel's intensity, then owing to progressive iteration processing, we can change the brightness values (the network status) in the picture in order to emphasize the edges in question. As soon as the processing is over, i.e. as soon as the shapes of the status do not alter or the changes are insignificant, we will get the status values between zero and one. The values close to zero represent the edges and the values close to one represent the area without edges. Provided we choose a fix threshold to which we compare these values, we'll obtain a final picture with sharply detectable edges. For the simplicity sake there is enough to set up the threshold to the value 0,5. The other values do not give mostly markedly distinct results.

As an example of utilizing the above mentioned continuous Hopfield's neural networks we can introduce the detection of edges in MR – picture that was shot by the tomograph in cooperation with DKFZ Heidelberg – see also Fig. 4a. The picture has a dimension 256×256 pixels, the brightness value are normalized to an interval $< 0;1 >$. In the Fig. 4 b till e are shown the individual phases of the normalization, i.e. different levels of edge detection. The pictures depict the situation (status) after 5000, 20000, 35000 and 50000 iterations.

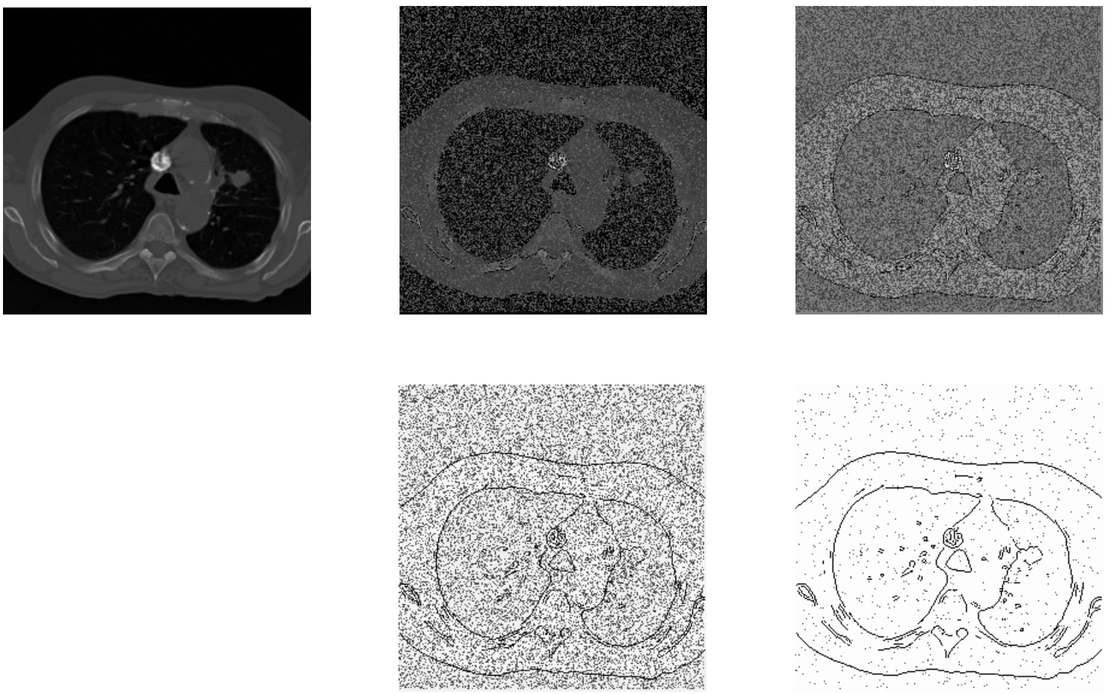


Fig. 4a-e. The individual picture's status during the detection of an edge. The first picture on the foregoing page is the final picture after the finalizing the interaction and shows the edges, that had been found.

The mentioned method shows good properties, but there is also a disadvantage: it is demanding as concerns memory and time.

2.2.3. Segmentation of MR pictures by utilizing Kohonen's map

The segmentation of the picture belongs to the basic operations during the preprocessing stage. Especially in the field of medical science is the utilizing of segmentation very large as it enables to detect remarkable areas. In this part of the lecture we will show a simple segmentation by means of Kohohen's neural network. We will come out from the specific configuration of Kohohen's grid and as the result we will get first of all the outline of the monitored area and, owing to secretiveness of this outline, also the area that is circumscribed by it (Reyes-Aldaroso, 2000).

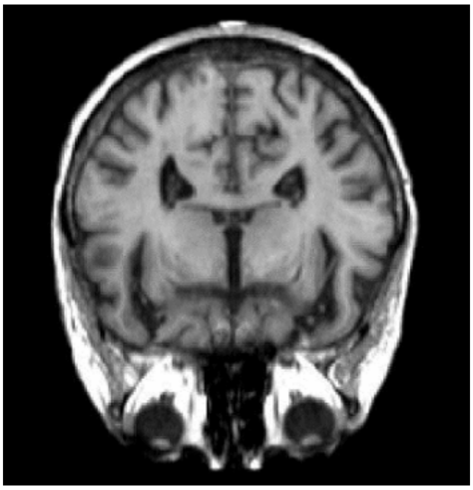
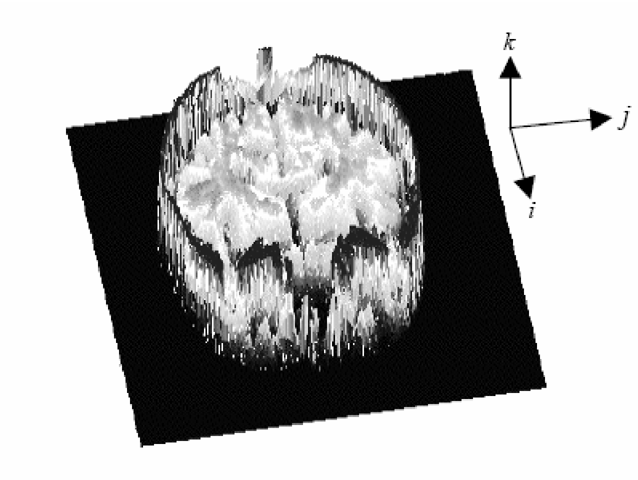


Fig. 5. MR image of the head.
Taken over from (Reyes-Aldasoro, 2000).



Fog. 6. Space figuration of the MR image from the Fig. 5.
Taken over from (Reyes-Aldasoro, 2000).

In the Fig. 5 is depicted MR – picture of the head, shown in transverse cross section. The picture is shown in grey tints and each the pixel value is depicted in the interval 0 - 256. The dimensions of the picture are 512×512 pixels. Each the pixel – value can be interpreted as the size and thus the picture can be depicted as 3D one (spatially, three dimensional) – see the Fig. 6. The same or similar values corresponds always to one certain type of the tissue and thus if we choose a threshold segmentation with two threshold values – the upper one and the lower one, we can select the respective tissue in question. In this way we will obtain a pixel picture that demarks the frame the area of interest.

The picture was preprocessed by the threshold method („thresholding“) for the value of the lower threshold - 145 and of the upper threshold – 210. Thus we gained the respective tissue corresponding to the bone of the skull, - see the Fig. 7.

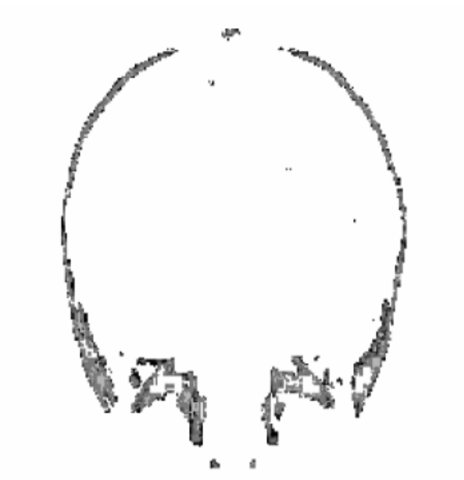


Fig. 7. Picture obtained by the threshold method of intensity from the Fig. 5, that is bigger than 145 and lesser than 210.
Taken over from (Reyes-Aldaroso, 2000).

Fig. 8. Final ascertained contour.
Taken over from (Reyes-Aldasoro, 2000).

The Kohonen's neural network consists of neurons that are interconnected by lateral connections in such a way so that they topologically create a closed structure, similar to a necklace. The assumption of the grid closure guarantees automatically, that the area - contours found here are closed, what used to be a serious problem in case of other classical methods of picture processing. The neural network has three entries: x – coordinate, y – coordinate and the intensity value in the certain, given spot, i.e. $f(x, y)$ from the thresholded picture. The training sample is created by group of three $\{x, y, f(x, y)\}$.

Particular values for training of the neural network were as follows: The number of neurons was chosen by experiment to 80 (the more of neurons, the better resolution thus the accuracy of the contour), the parameter of training was set up to 0,2 and during the training was linearly diminished to zero. The number of training was chosen to a number 15000. The neighborhood of the neuron was initially chosen at the size of 5 neurons (2 neurons to the either side of the middle neuron) and step by step diminished to one neuron only. The weights (centers) were set up to random values. We recommend in this case to set up the initial values of weights (coordinates of neurons) analytically by means of polar coordinates, i.e. to generate the initial shape in a form of a circle. The gained result is shown in the Fig. 8, where is also shown the contour entered into the original thresholded MR – picture.

2.2.4. Detection of tumors by utilizing RBF – network

In this chapter we will present the example of detection (classification) of the cancer tumors in the pictures from tomograph. The aim of this chapter is to show in which way is it possible to adopt and to use the neural networks for the detection of tumors, see e.g. (Karkanis et al., 2000).

The tumor detection principle is as follows: In the gained pictures shall an expert – a doctor specify areas that correspond to the cancer tumors. This set of pictures shall serve after pre-processing phase for practicing of the neural network. On the basis of the pictures are being calculated various characteristics (symptoms), that shall serve for practicing of the neural network. The symptoms should describe in the best way the different properties of a healthy tissue and the tissue hit by the tumor. Moreover, it is necessary to reduce the number of accessible data, in other words – it is not possible to let enter at the gate of the neural network all the pixels of the picture. As soon as the network is trained, it shall be possible to use it for detection of tumors in new pictures.

The procedure itself is as follows: First, we set up so called *co – occurrence matrices*. These shall aggregate the space dependences of the individual pixels in the picture. Consequently we use these matrices for computation of various statistics that depict the important properties and simplifies the data into a form of few digits. As this is classification task, it is necessary to specify during the practicing the respective outputs. These outputs are binary ones, e.g. the sign zero („0“) is the tissue without tumor and the sign one („1“) corresponds to the tumor tissue. Consequently, we set up from the computed signs the symptom's vectors. These vectors shall create practicing, pertinently validation and testing set, that shall be used for practicing of artificial neural network, in our case of the RBF – network.

As soon as we set up the co-occurrence matrices, we use them for computation of various statistics, specifically for energy computation (the second angle moment), correlation, inversion difference moment and entropy.

The detection of the tumor on the picture of the woman's breast with cancer tumor made by tomograph is shown in the Fig. 9. This picture was acquired in cooperation with DKFZ Heidelberg. In Fig. 10 the picture of the requested output is depicted.

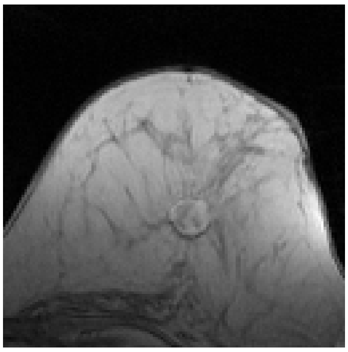


Fig. 9. Transverse cross-section of the tissue with cancer tumor.

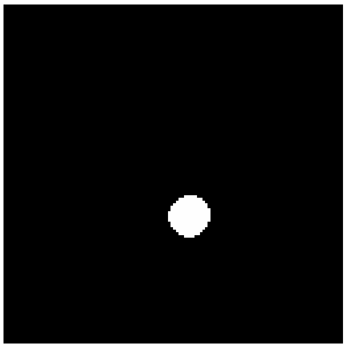


Fig. 10. The picture of the requested output. This binary picture is a mask to identify the tumor occurrence and serves as a correct output (the teacher) for training of neural network.

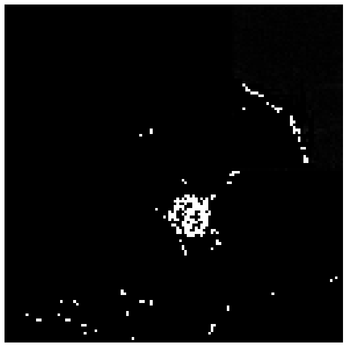


Fig. 11. Output from the network during the testing the learned network on the picture shown above.

As this is the case of typical classification task and we want to demonstrate here the use of RBF – network, we have chosen just this network. The number of entries is $4 \times 4 = 16$, as we count the above mentioned numerical statistics the for each matrix of the mutual occurrence. The number of neurons in the latent layer was, based on the experiments, set up to 165. The activation functions of these neurons are the Gauss – functions. The widths were different and they were chosen according to the distance to the nearest sample. In the outlet layer was one neuron with identity as activation function.

In Fig. 11 is shown a real output from the neural network after learning and verification 1 on the original picture[1].

Conclusion

The artificial neural works as mathematical algorithm can be successfully utilized in data processing. The very valued is the capability to learn form the examples and to provide the correct results even in the case of unknown samples. Due to the fact, that the neural networks are able to create their own intrinsic models, sometimes based on very vague information in data, they are suitable just for this area of biomedical engineering.

When utilizing the artificial neural networks in the biomedical engineering, there comes up the question of appropriateness and especially trustworthiness of their use. This is why we paid a higher attention particularly to the preparation and preprocessing of data and consequent suitable approach to the evaluation of acquired results.

[1]We would like to remark, that we demonstrate here the process on the same picture, but in fact the network was trained on the number of different pictures we had at disposal and the here presented picture was not used for training, but for testing purposes so that there was possible to demonstrate the progress on one picture and to have the results relevant.

Bibliography

Andersson A., Davidsson P., Lindén J.: *Measure-based classifier performance evaluation*, Pattern Recognition Letters, Elsevier, Vol. 20, pp. 1165-1173, 1999

Breiman L.: *Bagging predictors*, Machine Learning, Vol. 24, No. 2, pp. 123-140, 1996

Dietterich T. G.: *Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms*, Neural Computation, Vol. 10, No. 7, pp. 1985-1924, 1997

Dowla F. U., Rogers L. L.: *Solving Problems in Environmental Engineering and Geosciences With Artificial Neural Networks*, MIT Press, Cambridge, 1995

Dybowski R.: *Neural Computation in Medicine: Perspective and Prospects*, Proceedings of the Artificial Neural Networks in Medicine and Biology (ANNIMAB-1) Conference, Springer, pp. 26-36, Göteborg, 2000

Efron B, Tibshirani R. J.: *An introduction to the bootstrap*, Chapman and Hall, New York, USA, 1993

Flexer A.: *Statistical Evaluation of Neural Network Experiments: Minimum Requirements and Current Practice*, Technical Report oefai-tr-95-16, 1994

Hecht-Nielsen R.: *Neurocomputing*, Addison-Wesley, California, 1990

Hopfield J. J., Tank D. W.: *„Neural" computation of decisions in optimization problems*, Biological Cybernetics, Vol. 52, pp. 141-152, 1985

Chao C. H., Dhawan A. P.: *Edge detection using a Hopfield neural network*, Optical Eng., Vol. 33, No. 11, 1994

Jakob M.: *Symbolic Rule Extraction using Artificial Neural Networks*, thesis, FEL ČVUT, 2003

Jiřina M., Jiřina M., ml.: *Boundary Effect in Multivariate Data*, GESTS International Transaction on Computer Science and Engineering, Vol. 11, No. 1, pp. 85-94, ISSN 1738-6438, 2005 (a)

Jiřina M., Jiřina M., ml.: *How does the space where e-Golem walks looks like*, Interdisciplinary Aspects of Human-Machine Co-existence and Co-operation, pp. 271-279, Publisher ČVUT, Praha, 2005 (b)

Jiřina M., Smrčka P., Trefný Z., Hána K.: *New Robust Methods for Detection of Systolic Complexes in the Signal Acquired from Quantitative Seismocardiograph*, 6 p., CD Proceedings, EMBEC 2005 (c).

Karkanis S. A., Iakovidis D. K., Maroulis D. E., Magoulas G. D., Theofanous N. G.: Tumor *Recognition in Endoscopic Video Images using Artificial Neural Network Architectures*, EUROMICRO, IEEE, 2000

Kononenko I., Bratko I.: *Information-Based Evaluation Criterion for Classifier's Performance*, Machine Learning, 6, pp. 67-80, Kluwer Academic Publishers, Boston, 1991

Kotek Z., Vysoký P., Zdráhal Z.: *Kybernetika*, SNTL, Praha, 1990

Lavesson N.: *Evaluation of classifier performance and the impact of learning algorithm parameters*, Master Thesis No. MSE-2003:18, Dept. of Software Eng. and Comp. Science, Blekinge Institute of Technology, Sweden, 2003

Masters T.: *Practical Neural Network Recipes in C++*, Academic Press, Boston, 1994

Metz Ch. E.: *Basic Principles of ROC Analysis*, Seminars in Nuclear Medicine, Vol. 8, No. 4, pp. 283-298, 1978

Micheli-Tzanakou E.: *Supervised and unsupervised pattern recognition: Feature extraction and computational intelligence*, CRC Press, Washington, 2002

Michie D., Spiegelhalter D. J., Taylor C. C (editors): *Machine Learning, Neural and Statistical Classification*, Ellis Horwood, Chichester, England, 1994

Mosteller F., Tukey J. W.: *Data analysis and regression – a second course in statistics*, Addison-Wesley, 1977

Ohno-Machado L., Musen M. A.: *Hierarchical Neural Networks for Partial Diagnosis in Medicine*, World Congress on Neural Networks, San Diego, Vol. 1, pp. 291-296, 1994

Ohno-Machado L., Walker M. G., Musen M. A.: *Hierarchical Neural Networks for Survival Analysis*, The Eighth World Congress on Medical Informatics MEDINFO 95, Vancouver, Canada, pp. 828-832, 1995

Partridge D., Abidi S. S. R., Goh A.: *Neural Network Applications in Medicine*, Proceedings of National Conference on Research and Development in Computer Science and its Applications (REDECS'96), University Pertanian Malaysia, Kuala Lumpur, pp. 20-23, 1996

Passold F., Ojeda R. G., Mur J.: *Hybrid Expert System in Anesthesiology for Critical Patients*, Proceedings of the 8th IEEE Mediterranean Electrotechnical Conference (MELECON'96), Italy, Vol. 3, pp. 1486-1489, 1996

Reyes-Aldasoro C. C.: *Image Segmentation with Kohonen Neural Network Self-Organising Maps*, 2000, citeseer.ifi.unizh.ch/460754.html

Rizzo D. M., Dougherty D. E.: *Characterization of aquifer properties using artificial neural networks: neural kriging*, Water Res., Vol. 30, No. 2, pp. 483-497, 1994

Salzberg S. L.: *On Comparing Classifiers: Pitfalls to Avoid and a Recommended Approach, Data Mining and Knowledge Discovery*, Kluwer Academic Publishers, Vol. 1, pp. 317-327, Boston, 1997

Stein R.: *Selecting data for neural networks*, AI Expert 2, pp. 42-47, 1993

Swingler K.: *Applying Neural Networks: A Practical Guide*, Academic Press, New York, 1996

Curriculum Vitae

Marcel Jiřina, PhD

Born: February 11th, 1972 in Prague

Education: MSc, (Ing., with honor) in Technical cybernetics, Faculty of Electrical Engineering, Czech Technical University in Prague, 1995
PhD, (Ph.D.) in Informatics and Computer Art, Faculty of Electrical Engineering, Czech Technical University in Prague 1999
MSc, (Mgr., with honor) in Software engineering, Faculty of Mathematics and Physics, Charles University, 2000
Dr, (RNDr.) in Software engineering, Faculty of Mathematics and Physics, Charles University, 2003

Experience: 1992 - 1995 Supporting science worker, Department of Control, FEE, Czech Technical University in Prague
1994 Practice, University of Sheffield, GB
1996 – 1998 Scientific and research worker, ICS, Academy of Science, CZ.
1998 – 1999 Compulsory Military Service
1997 – 2002 Scientific and research worker, Gerstner Laboratory for intelligent decisions and management, FEE, Czech Technical University in Prague
2002 – 2005 Scientific and research worker, Center of applied cybernetics, FEE, Czech Technical University in Prague
Since 2005 Assistant professor, FBME, Czech Technical University in Prague, Deputy Chief of Department of Biomedicine informatics, Vice-dean for pedagogical activities, Member of Scientific Council FBMI

Pedagogical activity:

Introduction to systems and signals, 2+2, credit inclusion, examination, compulsory, attendance, FBME, Czech Technical University in Prague (since 2005)
Neural network, 2+0, examination, optional, attendance, FSV IES UK (Charles University)
Softcomputing, 2+0, examination, optional, attendance, FEE, Czech Technical University in Prague (collaborator on lectures and tutorial activity) (2003 – 2004)
Cognitive processes, 2+2, credit inclusion, examination, compulsory, optional, FEE, Czech Technical University in Prague (collaborator on lectures and tutorial activity) (2003 – 2004)
Alternative Approaches to Economic Modeling, 2+0, credit inclusion, compulsory-optional, FSS IES Charles University (2004 – 2006).