Synergies between Statistical Data Analysis and Neural Networks in the Control of Rotary Blood Pumps

Georg Dorffner, Christian Stöcklmayer
Dept. of Medical Cybernetics and Artificial Intelligence
University of Vienna; Freyung 6/2, A-1010 Vienna, Austria
and Austrian Research Institute for Artificial Intelligence
georg@ai.univie.ac.at

Christian Schmidt, Heinrich Schima
Dept. of Cardiothoracic Surgery, LBI-CBR
University of Vienna, Währinger Gürtel 18-20, A-1090 Vienna

Abstract

In this paper we report about the application of multilayer perceptrons to three important tasks in the control of rotary blood pumps, namely the estimation of left atrium pressure, and the indication of suction, as well as danger of suction, in the left atrium. Special focus is laid on the value of traditional techniques for statistical data analysis such as principal component analysis (PCA) as tools to guide the design of the neural network solution. Eleven parameters derived from actual measurements during the performance of the pump were available as input for the three tasks. With the help of PCA, they could be reduced to three major components and thus visualized. This visualisation served as guidance for the proper choice of network, and as a tool for initialization to improve learning.

Category: Applications

Keywords: principal component analysis, multilayer perceptrons, initialization, medical application, rotary blood pumps, control

Preferred presentation: oral; poster ok

1 Rotary blood pumps

Rotary blood pumps are used as heart assist systems, where they pump the blood via an inflow cannula from the left atrium of the heart through the device and an outflow cannula into the aorta [Schima et al. 1991]. Such pumps create a pressure head between inlet and outlet, which depends mainly on pump speed and flow, but which is independent from the absolute pressure values at the inlet and outlet. Therefore rotary blood pumps can exert a large negative pressure on the atrium ("suction") with deleterious effects on the atrial wall, the blood and the lung.

At the bioengineering group of the Department of Cardiothoracic Surgery at the University of Vienna, a controler for such a blood pump has been developed ([Schima et al. 1992]). Currently, improvements – especially with respect to adaptivity – are under development. One main problem is to detect suction to determine necessary reductions of the blood flow. For this problem, a "suction detector" indicating either the presence or the danger of suction would be of prime import. It would greatly simplify the monitoring of the patient and could be included in an adaptive control system. Another problem of great interest is the estimation of the left atrium pressure (LAP) – another indicator for critical situations –, which could only be measured with additional invasive devices and their inherent risks.

The aim of the project reported here is to evaluate neural networks with respect to their ability to form a suction detector and/or LAP estimator. The input are 11 characteristic values (such as signal mean, variance and range) computed from three measurable signals, namely pump speed, pump flow, and aortic pressure. Data from a mock simulation (water model) were taken in sequences, where always one of the measurable parameters was varied to cover an as large range of system states as possible. Thus a total of 1181 data sets were derived, which were split about in half for training (590) and cross-evaluation (591).

2 Principal component analysis for data evaluation

According to this brief description, the goal of this project was to develop solutions for three tasks – one approximation task (the estimation of LAP), and two classification tasks (detection of "suction" and "danger of suction"). In order to be able to properly decide about the type of neural network employed we performed a principal component analysis on the training set of 590 cases, each of which being an 11-dimensional vector. This analysis revealed that the first three principal components cover about 83 % of the total variance in the data – enough to be able to visualize the data distribution in a 3-dimensional plot. It should be noted that this does not follow from the fact that three measurable signals were the basis for the data, since the 11 characteristics are largely independent from each other. Figure 1 depicts all data points of the training set, distinguishing between cases of "suction" (grey points) and "no suction" (black). Figure 2 is a plot using only the "no suction"-data, distinguishing between "danger of suction" (grey) and "no danger."

With respect to the three tasks, these plots nicely supported the choice of approach toward a solution. The classification of "suction" vs. "no suction" appears almost perfectly linearly separable (figure 1), in fact suggesting a solution using a simple linear discriminator along the main axis of the distribution. The classification of

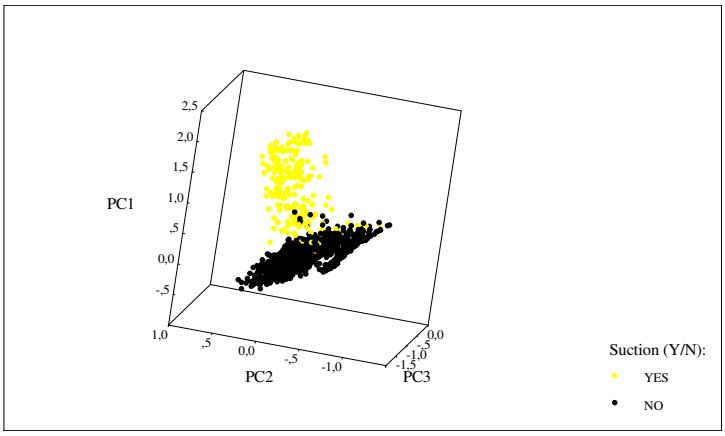

Figure 1: The distribution of the whole training set, with a distinction between "suction" (grey dots) and "no suction" (black)

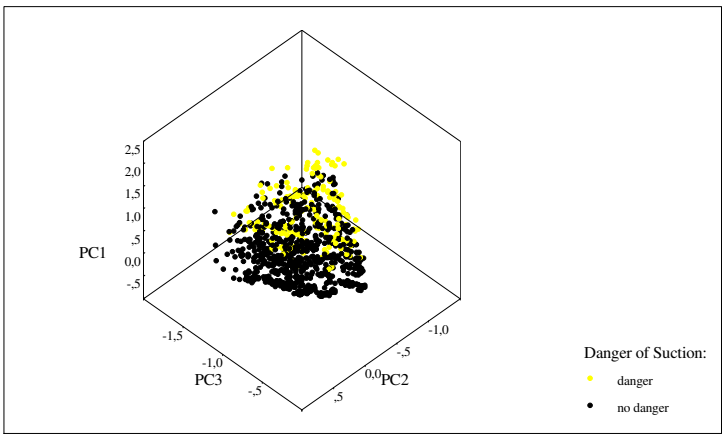

Figure 2: The distribution of the training data in the "no suction" region, with a distinction between "danger of suction" (grey) and "no danger" (black)

"danger of suction" vs. "no danger" appears to be separable to a certain extent, but with non-linear decison boundaries. Furthermore, the data distribution suggests a convenient way of initializing input-to-hidden weights of a multilayer perceptron (see below). The third task – the estimation of LAP – revealed, through coloring according to several ranges of LAP (not depicted), a distribution with no obvious structure or decision regions. These observations led to the following solutions.

3 Neural networks employed

As suggested above, the task of distinguishing "suction" from "no suction" was not approached by a neural network but by implementation of a simple linear discriminator. The first three principal components helped in finding the major axis of distinction, as well as the threshold for separation (figure 1).

The estimation of LAP was approached by a regular three-layer perceptron (i.e. one hidden layer) using backpropagation. The network (optimized empirically in several independent training runs) consisted of 11 input units, 10 hidden units, and 1 output unit. Weights were initialized in the range [-0.1,0.1]. During training, learning rate (η) and momentum rate (α) were changed according to the following schedule: 200,000 training steps with $\eta = 0.8, \alpha = 0.9$; 200,000 training steps with $\eta = 0.7, \alpha = 0.8$; 100,000 steps with $\eta = 0.4, \alpha = 0.5$. This schedule was also developed empirically. One training step consisted of a random draw on one input pattern and one cycle of update and backpropagation through the network.

The classification of "danger of suction" vs. "no danger" was also approached with a three-layer perceptron. Here, however, the visualization of the data could be exploited to initialize the hidden layer so as to form near-to-optimal decision regions before training. The technique used was initialization through Voronoi tesselation of positive and negative class samples, by setting input-to-hidden unit weights and bias θ of a unit j to

$$w_{ij} = x^{+}_{ji} - x^{-}_{ji} \tag{1}$$

$$\theta_j = \frac{1}{2}\sum_{i=1}^{n}({x^{+}_{ji}}^2 - {x^{-}_{ij}}^2) \tag{2}$$

where x^{+}_{ji} and x^{-}_{ji} are the i-th coordinate of the j-th positive and negative samples, respectively.

This method is a direct analogy to traditional classification theory and has been reported at several places ([Bedworth 1989, Smyth 1992, Dorffner & Porenta 1994]). To choose optimal positive and negative class samples, single data sequences, for which one of the system parameters was varied (leaving the other ones constant) until suction occured, were also depicted using the first three principal components. All of them had the characteristics of the one shown in figure 3, i.e., points moved nearly linear until during suction they broke off to a separate area in data space. For the classification problem, the last point before breaking off was defined as bearing the label "danger of suction." This point and the one immediately before were used as samples for a tesselation, for each of 10 randomly chosen data sequences.

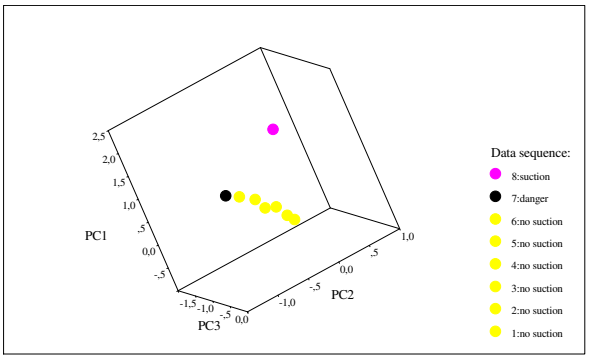

Figure 3: A single data sequence from no suction to suction

After initialization, two strategies are possible. Either, regular backpropagation is applied with the aim of fine-tuning input-to-hidden weights and to set hidden-to-output weights. If initialization is appropriate, smaller training times are expected (see [Smyth 1992]). Or input-to-hidden weights are kept fixed (under the assumption that they are sufficient for overall separation) and only hidden-to-output weights are trained using a simple delta rule.

The network consisted of 11 input, 10 hidden and 2 output units representing the two classes by 1/0 and 0/1 patterns. Two thresholds θ_{12} and θ_{21} were used to interpret the results. The first indicates by how much more the first unit must be activated to view the result as the 1/0 pattern. The second threshold does the same for the other class. By varying this threshold, sensitivity (percentage of correct positives) and specificity (percentage of correct negatives), and at the same time a third class of non-decidable inputs, can be hand-tuned to approach desired levels for actual application. For instance, during control it might be desirable to either recognize as many "danger" points as possible, while misclassifying as few as possible (and rather output "no decision").

For all training runs on the third task, a special training strategy with respect to data sampling had to be employed. Since "danger of suction" cases were only about one seventh of the total training set, training the network by epochs led to extreme overemphasis of negative cases. Thus, at frequent points in time during training, the network was evaluated (with the training set) and if the two classes were recognized differently the probability of picking a case of a certain class was increased for the under-represented one. This way, a balance between learning the two classes could be reached.

4 Results

The simple discriminator for detection of "suction" reached a performance of over 95 % sensitivity and 95 % specificity in the mock circulation.

Left atrium pressure (LAP) could be estimated by the multilayer perceptron with

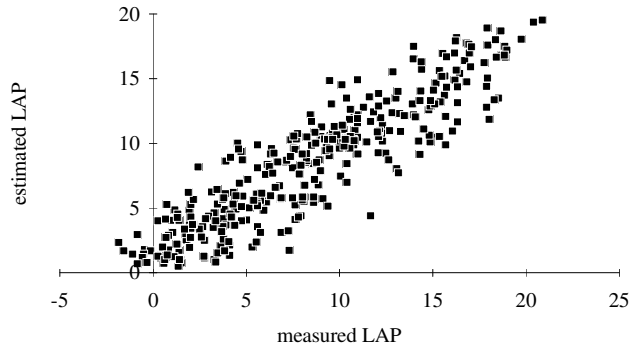

Figure 4: The regression diagram of estimated over real values of LAP

a performance of +/- 1.8 mm Hg, which is a reasonable estimate for the practical purposes of controling rotary blood pumps. Figure 4 shows the regression diagram of real and estimated value of the best result with a regression factor r^2 of 0.688 over the given test set of the in-vitro setup.

Finally, results for the detection of "danger of suction" are depicted in figure 5 for the initialized multilayer peceptron, where only hidden-to-output weights were trained. For comparison, in figure 6 results are shown for a network of the same size without initialization trained by regular backpropagation (with the same learning rate). The results are depicted as two curves of percentage of correctly classified "danger" (or "no danger," respectively) over percentage of misclassified cases of the other class, by varying the two thresholds θ_{12} and θ_{21} between 0 and 1. As an example, taking the top points of both curves in figure 5 yields about 75 % correctly classified "danger" points (i.e. sensitivity; x-axis of dotted curve), and about 72 % correctly classified "no danger" points (i.e. specificity; x-axis of black curve), while about 18 % of "no danger" are incorrectly recognized as "danger" (y-axis of dotted curve), and about 16 % of "danger" points are incorrectly classified as "no danger" (y-axis of black curve). The remaining percentages of cases would result in "no decision." The corresponding values of the regular backpropagation training are 76 % sensitivity and 79 % specificity. While these results are slightly better, the initialized network can be said to reach performance in the same range. Several such runs were performed yielding an average of around 48,000 training steps for backpropagation and around 28,000 for the initialized network.

5 Discussion

These results demonstrate that a prior statistical analysis of application data can reveal decisive information about the proper choice of neural network, as well as aid in simplifying the training process through initialization. Among the three tasks to be tackled, one could be solved with a rather simple solutions, avoiding neural networks where they are inappropriate. Multilayer perceptrons proved appropriate for the approximation task of estimating LAP, while the third task (classifying "danger of suction") could be solved in a much simplified manner as compared to "blind"

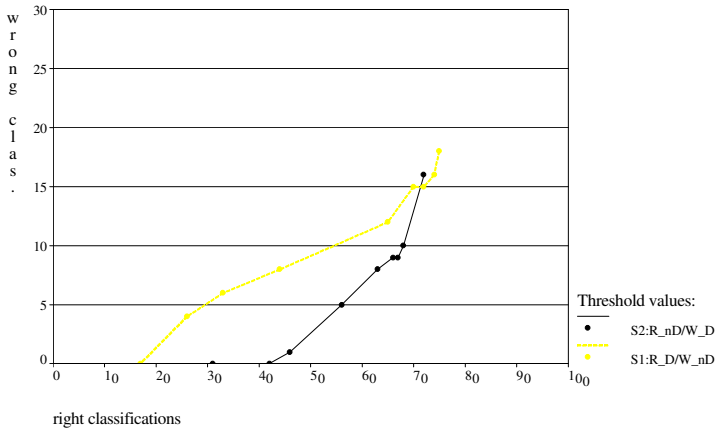

Figure 5: Curves of correctly classified classes over errors in the other class for "danger" (dotted curve) and "no danger" (black) for the initialized multilayer perceptron trained by delta rule

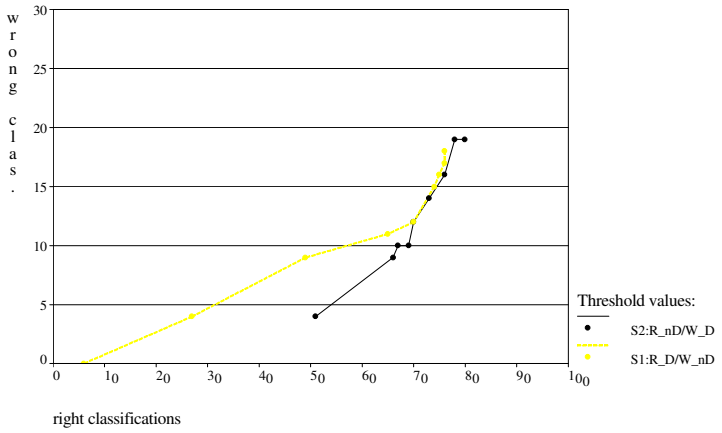

Figure 6: Curves of correctly classified classes over errors in the other class for "danger" (dotted curve) and "no danger" (black) for the multilayer perceptron trained by regular backpropagation

application of multilayer perceptrons. The comparison with a regular backpropagation run using the same parameters shows that initialization plus training of only hidden-to-output weights can lead to approximately the same performance, but with much fewer training steps (each of which is much less time-costly than a backpropagation step). Furthermore, this simplified training has a guaranteed solution (convergence theorem of the delta rule), provided the initialization led to sufficient separators. In this application this was the case (which also proved to be robust in comparative runs where different 10 data series were used for initialization).

With respect to the application, the results show that two of the tasks could be solved satisfactorily (LAP estimation and recognition of "suction"). The results of classifying "danger of suction," while helpful for supporting the pump control (the network for this task has been cross-evaluated in an in-vivo experiment), it will still have to be improved. One main reason why performance might not be at its possible maximum is that the available data was not collected originally with the goal of detecting "danger of suction" in mind. In other words, the resolution of data points in the crucial area (the areas where the hyperplanes are set) is very likely to be too coarse to permit good generalization. Secondly, the model does not take into account any sequential dependencies yet, which of course occur in a real setting. To explore these dynamics for improving the results will be the content of future research in this project. Furthermore, the transfer from the in-vitro study to the in-vivo experiment is still under investigation.

6 Acknowledgement

This project is supported by the Jubiläumsfonds der Oesterreichischen Nationalbank, under grant no. 4709.

References

[Bedworth 1989] Bedworth: Improving upon standard pattern classification algorithms by implementing them as multilayer perceptrons, RSRE Memorandum 4346, 1989.

[Dorffner & Porenta 1994] Dorffner G., Porenta G.: On Using Feedforward Neural Networks for Clinical Diagnostic Tasks, to appear in: *Artificial Intelligence in Medicine*, special issue on Neurocomputing in Medicine, 1994.

[Schima et al. 1991] Schima H., Thoma H., Wieselthaler G., Wolner E. (eds): *Proceedings of the international workshop on rotary blood pumps*, Vienna, Austria: Chirurgisch. Univ. Klinik, 1991.

[Schima et al. 1992] Schima H., Trubel W., Moritz A., Wieselthaler G., St"ohr H.G., Thoma H., Losert U., Wolner E.: Noninvasive Monitoring of Rotary Blood Pumps: Necessity, Possibilities, and Limitations, *Artificial Organs* 16(2)195-202, 1992.

[Smyth 1992] Smyth S.G.: Designing Multilayer Perceptrons from Nearest-Neighbor Systems, *IEEE Transactions on Neural Networks* 3(2)329-333, 1992.