About the Analysis of Septic Shock Patient Data

Jürgen Paetz[1,2], Fred Hamker[1,2], and Sven Thöne[2]

[1] J.W. Goethe-Universität Frankfurt am Main,
Fachbereich Informatik, AG Adaptive Systemarchitektur
Robert-Mayer-Straße 11-15, D-60054 Frankfurt am Main, Germany
[2] Klinikum der J.W. Goethe-Universität,
Klinik für Allgemein- und Gefäßchirurgie
Theodor-Stern-Kai 7, D-60590 Frankfurt am Main, Germany
paetz@cs.uni-frankfurt.de
http://www.medan.de

Abstract In intensive care medicine doctors are aware of a high mortality rate of septic shock patients. In this contribution we present the problems and the results of a retrospective, data driven analysis of two studies made in Frankfurt am Main between 1993 and 1997. Our approach includes the necessary steps of data mining, i.e. building up a data base, cleaning and preprocessing the data and finally choosing an adequate analysis for the medical patient data. We chose an architecture mainly based on a supervised neural network. The patient data is classified into two classes (survived and deceased). The importance of this classification for an early warning system is discussed.

1 Introduction – Medical and Data Background

In abdominal intensive care medicine patients are in a very bad condition. Often patients develop a *septic shock*, a phenomenon that is related to mechanisms of the immune system (see [1]) which is still a subject for research. The septic shock is associated with a high mortality rate of about 50%. It is always related to measurements out of the ordinary and often related to multiorgan failure. The exact medical definition for the septic shock and the epidemiology of 656 intensive care unit patients (47 with a septic shock, 25 of them deceased) is elaborated in a study made between November 1995 and December 1997 at the Klinikum der J.W. Goethe-Universität Frankfurt am Main [2]. The data of this study and another study made in the same clinic between November 1993 and November 1995 is the basis of our work. We set up a list of 140 variables, including readings (temperature, blood pressure, ...), drugs (dobutrex, dobutamin, ...) and therapy (diabetes, livercirrhosis, ...). – Our data base consists of 874 patients. 70 patients of all have a septic shock. 24 of the septic shock patients and 69 of all patients are deceased.

2 Preprocessing the Patient Data

Very important for medical data analysis, especially for retrospective evaluations is the preprocessing of the data. In *medical data mining*, after data-collection and

problem-definition, preprocessing is the third step.[1] Clearly, the quality of the results from data analysis strongly depends on the successful execution of the steps data-collection, problem-definition and preprocessing. The three steps are an interdisciplinary work from data analysts and doctors. The problems and our approaches to them are listed below:

1. We had medical data from two *different* studies. With the help of doctors we set up a common list of variables. Different units had to be adapted. Some variables are only measured in one of the two studies. It happened that time stamps were not clearly identifiable. Some data entries like *see above* or *zero* were not interpretable. So some database entries had to be ignored. The result is *one* common study with an unified relational database design including input- and output-programs and basic visualization programs.
2. *Typing errors* were detected by checking principal limit values of the variables. Blood pressure can not be 1200 (a missing decimal point). Typing errors in the date (03.12.96 instead of 30.12.96) were checked with the admission and the discharge day.
3. Naturally our medical data material is very *inhomogenous*, a fact that has to be emphasized. Each of the patients has a different period of time staying in the intensive care unit. For each patient a different number of variables (readings, drugs, therapies) is documented. So we had to select patients, variables and periods of time. Because different data are measured at different times of day with a different frequency (see table 1), which gave hard to interpretate multivariate time series, we used sampling-methods to get the measurements in regular 24 hours time intervals.

Table 1. Average of sampling rate of four measured variables from all patients without any preprocessing.

variable	days	hours	min
systolic blood pressure	1	12	11
temperature	1	12	31
thrombocytes	1	18	13
lactate	5	0	53

4. A lot of variables showed a high number of *missing values* (internally coded with -9999) caused by faults or simply by seldom measurements, see table 2 on the next page. The treatment of missing values in the analysis with neural networks is described in more detail in section 3.2.

In conclusion it is almost impossible to get 100% clean data from an enormous amount of different patient records. Nevertheless, we are sure that we have

[1] In a *prospective* evaluation the problem-definition phase takes place before the data-collection phase.

Table 2. Available measurements from septic shock patients after 24-hours sampling for six variables.

variable	measurements
systolic blood pressure	83.27 %
temperature	82.69 %
thrombocytes	73.60 %
inspiratorical O2-concentration	65.81 %
lactate	18.38 %
lipase	1.45 %

cleaned the data as good as possible with an enormous amount of time to allow analysis, see chapter 4.

3 The Concept of an Early Warning System

The first of two subsections presents the principal idea of an early warning system. The second subsection describes our neural network approach.

3.1 Principal Idea

The patients often change their conditions quickly. To assist the doctor to protect the patient's life, the main idea concerning the septic shock problem is not to make a prognosis about the survival of the patients, but to build up an *early warning system* to give individual warnings about the patient's critical condition to the doctor. The principle of such a system is shown in figure 1. In the periods of time U_i patients are uncritical, in K_j they are critical. The aim of an early warning system is to give an alarm as early as possible in the transition phases W_k $(k = 1, 3)$ and of course in K_j.

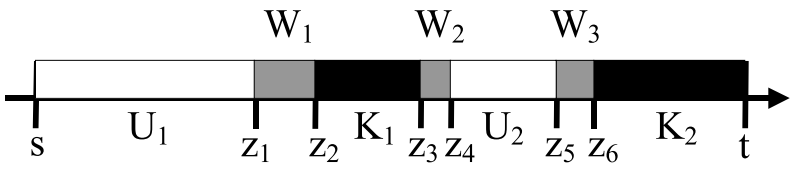

Figure 1. The concept of an early warning system: s = time of admission, t = time of death, $z_1, \ldots, z_6$ = change of state: begin and end, U_1, U_2 = uncritical period of time, K_1, K_2 = critical period of time and $W_1, \ldots, W_3$ = period of time: change of state.

To achieve knowledge about a patient being in a critical illness condition, we need to *classify* the vectors $(x_1, ..., x_n)^t$ composed of measurements or drugs x_i, $i = 1, \ldots, n$ with the outcome y_s (survived) resp. y_d (deceased).

Critical illness states are defined as those states which are located in areas of the input space showing a majority of measurements from deceased patients, see [8]. By detecting those states we expect to achieve a reliable warning, which should be as early as possible.

3.2 The Neural Network Approach

In the last years many authors contributed to machine learning, data mining, intelligent data analysis and neural networks in medicine (see [3] and [4]). Concerning our problem supervised neural networks have the following positive aspects: nonlinear classification, fault tolerance, learning from data and generalization ability. The aim of this contribution is not a comparison of statistical with neural network methods (see [5]) but to select an appropriate method that can easily be adapted to our data. Here, our aim is to detect critical illness states with a classification method. *Linear* classifiers did not seem to be suitable for a classification after having a look at the data. In addition, a nonlinear method surely detects a linear separability.

The neural network chosen for our classification is the supervised growing neural gas (abbr. SGNG, see [6]).[2] Compared to the multilayer perceptron, trained with backpropagation (see [9]), which has reached a wide public, this network achieved similar results on classsification tasks, see [7]. The results are presented in chapter 4.2. Its additional advantage is the ability to insert neurons within the learning process to adapt its structure to the data. The algorithm with its slight modifications and its parameters is noted down in detail in [8]. – It is based on the idea of radial basis functions (abbr. RBF, see [9]). The centers of the radial basis functions are connected through an additional graph that is adapted within the learning process. The graph structure allows to adapt not only the parameters (weights, radii) of the best matching neuron but also those of its neighbours (adjacent neurons).

One feature we additionally integrated into the SGNG algorithm is the toleration of the missing values within the adaptation and activation phase. If $x = (x_1, \ldots, x_n)^t$ is a n-dimensional data vector, you can project the vector x, so that no missing value is in the projected vector $x_p := (x_{i_1}, \ldots, x_{i_m})^t$, $\{x_{i_1}, \ldots, x_{i_m}\} \subset \{1, \ldots, n\}$, $m \leq n$, $x_{i_1}, \ldots, x_{i_m}$ are not missing values. Due to the fact that the SGNG is based only on distance calculations between vectors, it is possible to apply this standard projection argument to the adaptation and activation calculations of the SGNG, so that all calculations are done with the

[2] Logistic regression is a statistical alternative to supervised neural networks.

projected vectors x_p. Preliminary experiments showed that it is not appropriate to project to less than the half of the variables; sample vectors with more than 50% missing values are omitted. This procedure causes a statistical bias, but we believe that it is not high because the most part of the data is missing randomly.

4 Results

We give an impression of our results achieved up to now. We are aware of the problem that data from only 70 patients with a septic shock, including missing values in some variables, are not sufficient for *excellent* results but we are able to give some hints and first results in the right direction with the data available at the moment.

4.1 Basic Statistical Analysis

We calculated some statistical standard measures for each of the variables (mean, standard deviation etc.) including all patients or only the septic shock patients combined with all days or comprising only the last day of their stay in the intensive care unit. We detected some distinctions but they are usually not significant. Q-Q-plots show that the distributions are usually normal with an overlap of values from deceased and survived patients. A *correlation analysis* of the data shows high absolute values for the correlations between medicaments and variables, so surely the medicaments complicate the data analysis. Correlations between variables and {survived, deceased} are not high or not significant.

More interesting are the correlations $\widetilde{\mathrm{COR}}$ between variables X, Z, calculated one time with the sets X_d, Z_d of samples from deceased and one time with the sets X_s, Z_s of samples from survived patients and the corresponding differences taken from all patients and all days, listed in the table 3 on the next page. The significance level was calculated with SPSS 9.0. The correlations with significance level 0.01 are marked with an asterisk.

Both correlation values for the pairs urea, creatinin and arterial pO2, potassium are significant (level 0.01), so that the difference could be an indicator for survived or deceased patients. Therefore, these variables have to be measured very often to calculate the correlation in a time window during the patients actual stay at hospital. - In addition to the results in the section 4.2 an idea is to train a neural network with the correlation values to find out the exact threshold for a warning based on correlation values or combinations or modifications of such values (for first results see [8]). This seems to be reasonable because doctors reported that the *interdependence* of variables, measured from critical illness patients, could be disturbed.

Table 3. Correlations between two variables (all patients, all days of hospital stay) with the highest correlation differences ≥ 0.3 betweeen survived and deceased patients and frequency of measurement of each variable $\geq 20\%$. Significant correlations (level 0.01) are marked with an asterisk. GGT is the abbreviation of gammaglutamyltransferase.

variable X	variable Z	$\widetilde{\mathrm{COR}}(X_s, Z_s)$	$\widetilde{\mathrm{COR}}(X_d, Z_d)$	difference
inspir. O2-concentration	pH	-0.03	-0.39*	0.36
leukocytes	GGT	0.00	0.32*	0.32
iron (Fe)	GGT	0.31*	0.01	0.30
(total) bilirubin	urea	0.26*	-0.07	0.33
urea	creatinin	0.14*	0.57*	0.43
fibrinogen	creatinin in urine	0.05	-0.31*	0.36
arterial pO2	potassium (K)	-0.13*	0.18*	0.31
thromboplastintime	chloride	0.24*	-0.07	0.31

4.2 Classification of Septic Shock Patient Data

In this section we use the neural network architecture described in section 3.2. The doctors gave a recommendation which variables are the most important ones for a classification. The chosen variable set V is composed of: pO2 (arterial), pCO2 (arterial), pH, leukocytes, thromboplastintime, thrombocytes, lactate, creatinin, heart frequency, volume of urine, systolic blood pressure, frequency of artificial respiratory, inspiratorical O2-concentration, antithrombine III, dopamine and dobutrex. The variables are normalized (mean 0, standard deviation 1) for analysis. The *classification* is based on 2068 measurement vectors (16-dimensional samples) from variable set V taken from 70 septic shock patients. 348 are deleted because of too many missing values in the sample. With 75% of the 1720 remaining samples the SGNG was trained and with 25 % samples from *completely other* patients than in the training set it was tested.

Table 4. Correct classifications, sensitivity, specifity with standard deviation, minimum and maximum in % from three repetitions.

measure	mean value	standard deviation	minimum	maximum
correct classification	67.84	6.96	61.17	75.05
sensitivity	24.94	4.85	19.38	28.30
specifity	91.61	2.53	89.74	94.49

The network chosen was the one with the lowest error on the smoothed test error function. Three repetitions of the complete learning process with different, randomly selected divisions of the data were made. The results are presented in table 4. To achieve a *generally applicable* result ten repetitions would be better

but here it is already clear that with the low number of data samples the results can only have a *prototypical* character, even with more cleverly thought-out benchmark strategies. Some additional results are reported in [8]. On average

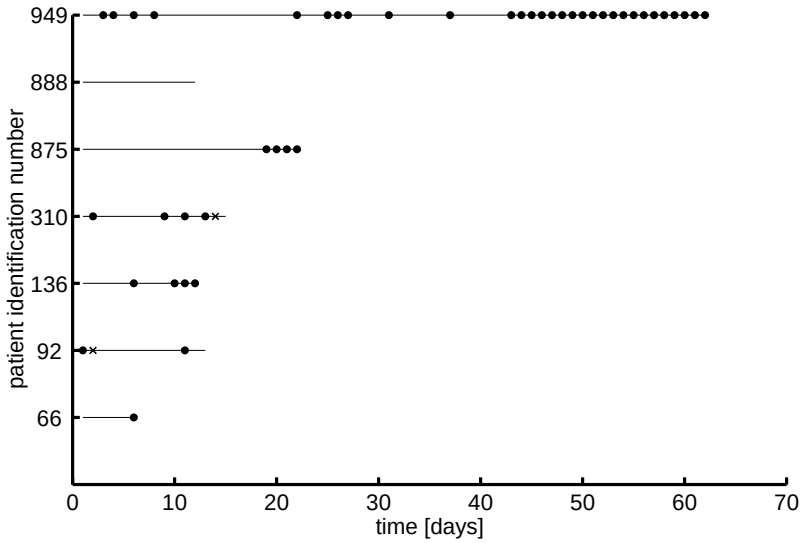

Figure 2. Deceased septic shock patients during their hospital stay with warnings (filled circles). A too high number of missing values causes some missing states (crosses). If there is no marking then no warning is given.

we have an alarm rate (= 1− specifity) of 8.39 % for survived patients showing also a critical state and a detection of about 1 out of 4 critical illness states. For such a complex problem it is a not too bad but clearly no excellent result. An explanation for this result are the different, individual measurements of each patient. In figure 2 the resulting warnings from classification are shown for 7 out of 24 deceased patients with a septic shock (to give an impression of the warnings over time). Not for each deceased patient exists a warning (patient with number 888) and some warnings are given too late (patient with number 66), i.e. the doctors knew already that the patient had become critical. So the ideal time to warn the doctor has not yet been found for all patients and it remains as future work.

5 Conclusion

We have presented a data analysis approach for the important medical problem *septic shock*. The results are basically encouraging for the doctors to achieve an early warning system for septic shock patients but the results are not final.

In spite of some severe restrictions of the data we succeeded in achieving some results by using several preprocessing steps. In the near future it is desirable to improve the performance of the results.

Further work will be a comparison of the achieved results with *scores*, that are known to have limitations in classifying individual patients (see [10]). Some results from *cluster analysis* are presented in [8]. To improve the results we are collecting data from septic shock patients from 166 clinics in Germany to evaluate our algorithms on this larger amount of patient data. Another approach is *adaptive rule generation* to explain the class boundaries in the data and at the same time to find out the necessary variables for the early warning system.

Acknowledgement: The work was done within the DFG-project MEDAN (Medical Data Analysis with Neural Networks). The authors like to thank all the participants of the MEDAN working group especially Prof. Hanisch and Dr. Brause and all other persons involved in the MEDAN project for supporting our work.

References

1. Fein, A.M. et al. (eds.): Sepsis and Multiorgan Failure, Williams & Wilkins Baltimore (1997)
2. Wade, S., Büssow, M., Hanisch, E.: Epidemiologie von SIRS, Sepsis und septischem Schock bei chirurgischen Intensivpatienten. Der Chirurg **69** (1998) 648–655
3. Lavrač: Machine Learning for Data Mining in Medicine. In: Horn, W. et al (eds.): AIMDM'99. LNAI 1620. Springer-Verlag Berlin Heidelberg (1999) 47-62
4. Brause, R.: Revolutionieren Neuronale Netze unsere Vorhersagefähigkeiten? Zentralblatt für Chirurgie **124** (1999) 692–698
5. Schumacher, M., Roßner, R., Vach, W.: Neural Networks and Logistic Regression: Part I. Computational Statistics & Data Analysis **21** (1996) 661–682
6. Fritzke, B.: Fast Learning with Incremental RBF Networks. Neural Processing Letters **1**(1) (1994) 2–5
7. Heinke, D., Hamker, F.: Comparing Neural Networks: A Benchmark on Growing Neural Gas, Growing Cell Structures, and Fuzzy ARTMAP. IEEE Transactions on Neural Networks **9**(6) (1998) 1279–1291
8. Hamker, F., Paetz, J., Thöne, S.: Detektion von kritischen Zuständen von Patienten mit der Diagnose „Septischer Schock" mit einem RBF-Netz. Interner Bericht, Fachbereich Informatik, J.W. Goethe-Universität Frankfurt am Main (2000)
9. Haykin, S.: Neural Networks: A Comprehensive Foundation. Prentice Hall (1999) 2^{nd} edition
10. Neugebauer, E., Lefering, R.: Scoresysteme und Datenbanken in der Intensivmedizin - Notwendigkeit und Grenzen. Intensivmedizin **33** (1996) 445–447