

About Adaptive State Knowledge Extraction for Septic Shock Mortality Prediction

Rüdiger W. Brause

J.W.Goethe-University, FB15, D-60054 Frankfurt a.M.

Email: Brause@CS.Uni-Frankfurt.de

Abstract

The early prediction of mortality is one of the unresolved tasks in intensive care medicine. This contribution models medical symptoms as observations caused by transitions between hidden markov states. Learning the underlying state transition probabilities results in a prediction probability success of about 91%.

The results are discussed and put in relation to the model used. Finally, the rationales for using the model are reflected: Are there states in the septic shock data?

1. Introduction

Medical care in intensive care units (ICU) has to deal with all consequences of operations and infection after operations. In 1992 a consensus conference [1] defined states of septic patient like the systemic inflammatory response syndrome (SIRS), sepsis, severe sepsis and septic shock. Although a dynamic modeling of septic patient states based on these definitions is very rare, it is very important. Approximately 60% of the patients acquire SIRS. Combined with an infection this leads to a sepsis in half of the cases. A septic shock is only developed in 7% of the cases, but 50% of the septic shock patients die.

Up to now, there are neither a successful clinical therapy to deal with this problem nor reliable early warning criteria to avoid such a situation. The event of sepsis and septic shock is rare and therefore statistically not well represented. Due to this fact, neither physicians can develop well grounded experience in this subject nor a statistical basis for this does exist. The diagnosis of septic shock is still made too late, because at present there are no adequate tools to predict the progression of sepsis to septic shock. No diagnosis of septic shock can be made before organ dysfunction is manifest.

The criteria for SIRS are both non-specific and potentially restrictive. Experience with the consensus conference definitions in clinical trials has highlighted the fact that they are unable to accurately identify patients with septic shock who might respond to interventions targeted to bacterial infections and its consequences, identify patients at risk for septic shock and to improve the early diagnosis

of septic shock.

Some years ago, Rangel-Fraustro et al.[5] treated the transition between septic states in the ICU by a Markov model, i.e. by a process where each transition is subject only to inherent transition probabilities, not determined by any further history. The transition probability matrix containing a strong diagonal was obtained using data of 2527 patients. The rationale behind is the idea that any therapy of sepsis have to be compared to the traditional therapy situation by its attached transition probabilities from a septic state to a non-septic one.

This idea might work well for groups of patients, but it allows no prediction and conclusion for a concrete case which can be seen as a probabilistic instance of the group. Our goal is the probability-based diagnosis of the individual case history. Therefore, we take the approach of modeling the septic state transition chain by hidden markov model (HMM) chains [4]. This approach has several advantages:

- ◆ The observed symptoms will be only regarded as probabilistic expressions of hidden, not known states. Therefore, the same symptom might serve either as indication for a good or a bad prognosis depending on the symptom history.
- ◆ The number of states themselves can be learned. This allows us to decide whether the states defined by the consensus conference are well chosen or not.
- ◆ All probabilities in the model will be obtained only by training with examples. This omits subjective definitions based on non-representative experience.

Now, let us introduce our model in the next section.

2. The model

The septic shock is assumed to occur after some prior states of the patient. In Fig. 1 a popular state transition model is shown, adapted from the model of [5].

The arcs are drawn according to an estimated transition probability. Since the transitions may occur rapidly, not all states are observed in all cases.

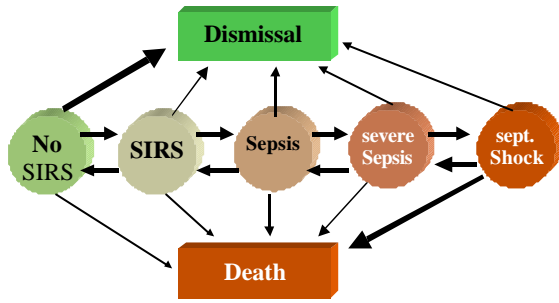


Fig. 1 States and transitions in ICU care

In our HMM, each patient class (surviving or lethal) defines its own transition probabilities between the states, especially to the death and dismissal state. Therefore, we distinguish between at least two hidden markov models: one for the surviving patients and one for lethal exit. Our diagnostic approach consists of presenting the patient data to a system which computes the probability for the data of a patient to be either part of the surviving or the mortal HMM.

This approach imitates the classic application of HMM in speech recognition where we have one HMM for each word, composed of the sequence of several phonemes. For each observed speech section and each particular HMM the probability is computed that the speech section has been produced by the HMM. The HMM with the highest probability is classified to represent the word which was spoken. This speech classification process can be adapted to our problem of patient classification. In our case, each phoneme sequence correspond to a patient record of symptoms. In principle, we only have two words i.e. two classes: survived and deceased.

3. Diagnosing the data

The performance of the system depends on the number of states, transition and symptom output probabilities for the two HMM models. Therefore, in order to fit clinical reality, we used an adaptive approach to extract the necessary knowledge for adjusting the corresponding parameters. In speech analysis, the well known Viterbi algorithm is used to adapt HMM of phoneme chains to speech reality [4]. By mapping the problem of septic shock state transition identification to phoneme transition we could use the adaption algorithms and were even able to partially use standard HMM speech recognition software [8] for patient data diagnosis.

Standard data analysis [2] showed that for septic shock data the symptoms for both patient groups are difficult to distinguish. In Fig. 2 two histograms of observations are shown.

Clearly, the high overlap of the Gaussian distributions of the observed variables makes a reliable diagnosis very difficult.

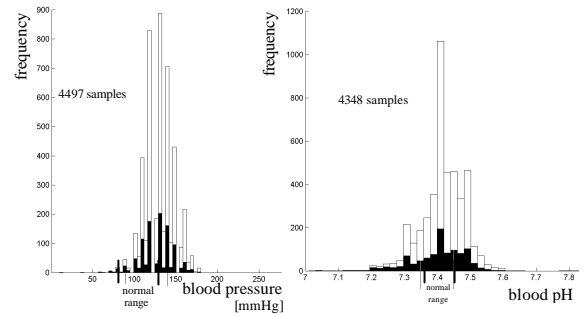


Fig. 2 Histograms for a) systolic blood pressure and b) pH value for survived (white boxes) and deceased patients (black boxes).

The question arises: How sensitive to such problems is the HMM approach?

3.1 Testing the diagnostic tool

Now, for our purpose we define a simplified version of the scenario presented in Fig. 1, because for the small number of septic shock patients the number of states and transitions should be as small as possible in order to avoid the “curse of dimensionality”[3]. In Fig. 3 the simplified HMM for testing is shown.

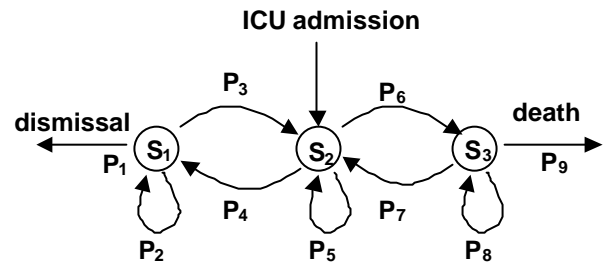


Fig. 3 The simplified HMM for testing

The model consists of three states: normal (S_1), critical (S_2) and very critical (S_3). For ICU admission we assumed the patient to be in an critical state, leaving the ICU either on dismissal or after death.

For the test data, we generated random sequences by accessing the state transition model shown in Fig. 3. The transition probabilities were assumed to be $P_1=0.1$, $P_2=0.4$, $P_3=P_4=0.5$, $P_5=0.3$, $P_6=0.2$, $P_7=0.4$, $P_8=0.5$, $P_9=0.1$. First, for each synthetic patient record we generated a random walk through the state graph until it exits and stored the resulting class, survived or deceased, in the record.

For each transition we assumed a symptom produced out of three possible Gaussian distributions of normal, critical and very critical symptoms $N_n(\mu_n, \sigma)$, $N_c(\mu_c, \sigma)$ and $N_v(\mu_v, \sigma)$. Since we can not observe the states directly, the symptoms are not equivalent to the states.

If the two patient classes {survived, deceased} would have equal probabilities for producing normal, critical or very critical symptoms we had no possibility for diagnosing them. Therefore, differences in the parameters for symptom producing were assumed for the classes. For instance, lethal patients produce in state S_3 a symptom from N_v with probability one whereas dismissed patients produce in this state only symptoms of N_c . In S_2 , we assumed $P(N_n | survived) = 0.1$, $P(N_c | survived) = 0.7$, $P(N_v | survived) = 0.2$, $P(N_n | deceased) = 0.1$, $P(N_c | deceased) = 0.45 = P(N_v | deceased)$. In state S_1 all classes produce symptoms of N_n with probability one.

In the general case we have to assume different means and variances of the three probability density functions (pdfs) individually for each transition. This results in too many parameters which can not be learned by our few medical training data. Thus, we reduced the complexity of the model by assuming only three fixed pdfs but state-dependent conditional probabilities.

Now, how does the overlap of the symptom producing probabilities influences the recognition probability? The overlap of the pdfs can be characterized by the common area of pdf intersection divided by the full pdf area under the pdf curves. This gives an overlap degree between 0 and 1 of the symptom pdfs, reaching 1 for full overlap and 0 in the limit of complete disparity. In Fig. 4 the distance between the centers of the Gaussian pdfs as function of the overlap degree is shown.

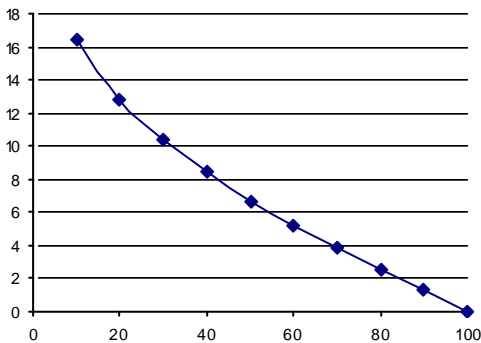


Fig. 4 The distance between the symptom distribution expectation values μ as function of the pdf overlap (in %). All variances σ are equal to 5.

The number of variables per symptom varied: we produced data sets with $n=2, 5$ and 10 variables.

The diagnosis was made on the basis of a training by the synthetic patient data. In order to learn the proper model, we did not assume a fixed number of states for the HMM but trained 4 models with 2,3,4 and 5 states and 1,2,3 and 4 symptom pdfs each, i.e. 16 different models in parallel. We repeated the process 20 times. For each training we chose 50% of the data for training and 50% for testing. In each data set we have 32.5 % deceased patients i.e. 65 of 200 and 130 of 400 patients.

For each test (diagnostic) sample the one HMM was chosen which obtained its diagnose with the highest probability. The classification probability of this best HMM as function of the symptom pdf overlap for the two classes using 400 patient records is shown in Fig. 5. We see that the diagnose of the HMMs is very good unless the symptoms become very similar.

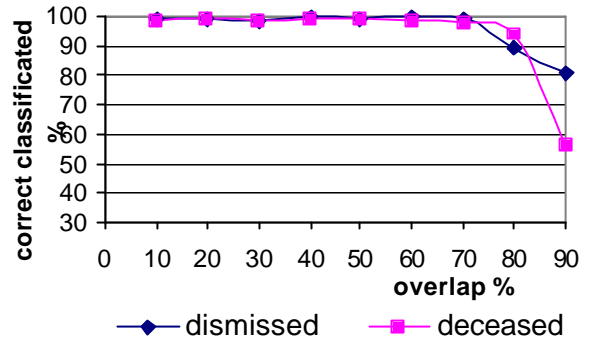


Fig. 5 The classification probability for the two classes as function of the pdf overlap (in %) for 400 patient records and 10 variables.

If we reduce the number of patient records, the performance drops slightly down in all overlap regions. Interestingly, reducing the number of variables per sample decreases also the classification performance. For illustration, in Fig. 6 the classification probability for the two classes is shown for the case of 100 different patient records and 2 variables.

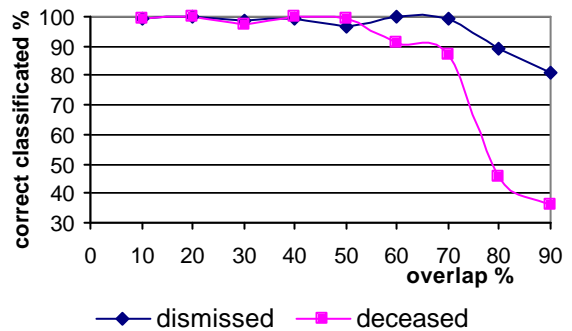


Fig. 6 The classification probability for the two classes as function of the pdf overlap (in %) for 100 different patient records and 2 variables

We observe also that the smaller number of training data is reflected in a higher variance of the classification probability.

3.2 Diagnosing patient data

We used the data of 32 survived and 38 deceased patients with septic shock collected in the multi-center study MEDAN [9]. Prior to diagnosis, the data were decorrelated. Thus, the HMM learning algorithm avoids learning all correlation matrix entries and concentrates on the main means and variances.

For a high number of dimensions the number of samples is too low and we run into the “curse of dimensionality”, see [3]. Therefore, we have to drop a number of variables. For the 108 recorded variables those with more than 16% missing values were dropped. This resulted in a set H_1 of 10 frequent variables not counting five dependent score variables. The three most frequent variables compose the set $H_2 \subset H_1$. Another criterion is quality: There is another subset H_3 consisting of 13 “medically important” variables. The intersection of the two sets is $H_4 = H_1 \cap H_3$ which contains only 5 variables.

Since we had only a small sample size, we used the leave-one-out method for training and testing: We used all records for training except one which was used for testing. This was repeated for all samples; the final test result is the average of all particular test results. The main test results after training are listed in Table 1, ordered by the number of variables used.

set	#var	dismissed %	deceased %
H_2	3	83.1	91.1
H_4	5	70.5	75.8
H_1	10	61.6	80.0
H_3	13	83.1	89.0

Table 1 The probabilities of a correct diagnosis

The training sets are very different and the results can hardly be compared. The increased test performance with the number of data, i.e. samples and variables, can not be observed generally since our results are influenced by both the curse of dimensionality and the training facility. In contrast to this, we note that the diagnostic performance for the real patient data depends also on the choice of variables. The best performance is given for the most frequent 3 variables.

4. Discussion

Using the hidden markov modeling we obtain a satisfying performance in diagnosing septic shock patients. Therefore, we might conclude that using hidden markov models is a good tool for extracting knowledge from patient symptom time series. In a limited sense, this is the case, but there is much more work to be done.

- The classification was based on the whole patient

record, i.e. the whole time sequence of symptoms. In clinical use, this is not available: Our task consists of giving early alarm in critical cases. This is only possible if the diagnostic system can also diagnose a part of the patient record, the beginning. Here, additional work for selecting appropriate sequence lengths are necessary.

- The results are based only on a small amount of data since we concentrated on the rare cases of abdominal septic shock. Nevertheless, for reliable statistical results this is not sufficient: A greater number of cases have to be included which will be eventually the case after the MEDAN study had terminated.

There are also more fundamental flaws of this approach which should be discussed here. Let us start with the model itself.

4.1 What is the best hidden markov model?

The knowledge extraction process used 16 HMMs with 5 states maximally for training and diagnosis. Unfortunately, none of the HMMs can be qualified as “best model”, e.g. by classification of more than 50% of the samples neither for synthetic data nor for septic shock patient data. Why? The reason is very simple: each state transition occurs with a certain probability. To stay in one state (for instance in S_2) for t time ticks, the probability is $(P_5)^t$ decreasing exponentially with increasing time. This situation does not reflect reality properly. For instance, if there is an average time for staying in a state the probability to stay either shorter or longer becomes small. The associated probability distribution may be a Gaussian or Bernouille one, but not a decreasing exponential. Thus, in order to approximate this behavior, additional states have to be used in the training process. In order to get only one model for all data we have to change the fundamental properties of our model resulting an a non - hidden markov one.

This is not the only critic in our approach. More fundamental, our model is based on the assumption that the patients stay in different states. Is this assumption justified by reality?

4.2 Are there states in the septic disease?

One of the most interesting questions is whether there are really septic disease states. Certainly, in the western culture symptoms like raised body temperature are treated as indications of a human state called “disease” whereas other cultures do not distinguish between healthy and ill people but only between less or more healthy ones. Here, the western concept of health states is contrasted by a continuous view of health not as a qualitative state but as a quantitative variable. What is reality?

In the study of Rangel-Fraustro [5] the transition probabilities between the states are much smaller than the probabilities to remain in the states, e.g. $\text{prob}(\text{severe sepsis} \rightarrow \text{septic shock}) = 0.2$, $\text{prob}(\text{severe sepsis} \rightarrow \text{severe sepsis}) = 0.7$. Does this prove that there are septic states? Certainly not. To see this, let us consider an example. Imagine that there is one variable which is observed, for instance the temperature. This might behave either continuous (here: linear) or in sudden jumps between certain state values, see Fig. 7.

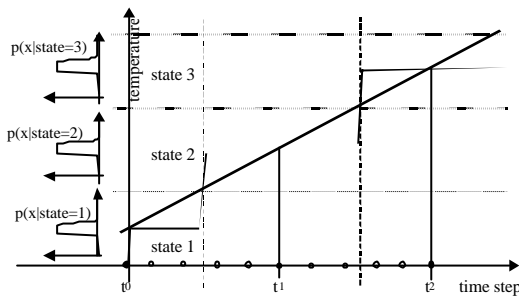


Fig. 7 The variable "temperature" and its sampled values

The temperature is sampled at certain time events, say t_0 , t_1 and t_2 , and classified into three states 1, 2 and 3. Now, if we count the number of transitions to the same state and to the next one, we have different states at different time points, i.e. a probability of one for the transition from state i to $i+1$. If we add more samples by subsampling (the hollow circles on the time line in Fig. 7) we have more samples for one state, in our example 5 samples instead of one. This gives a probability of 4/5 for remaining in one state and 1/5 for passing to the next one.

This example shows us two insights:

- ◆ The probability of a remaining in a state depends on the sampling rate, not on the fact whether there are states or not. The linear line without steps in Fig. 7 has the same probability of remaining in a state as the step function of real states. Therefore, the absolute values in the main diagonal of the probability transition matrix is no argument for the existence of states.
- ◆ The main difference between "states" and "non-states" is the nonlinearity in the probability distribution of the variables. The intuitive understanding of "state" implies a certain time of stay at the same value (or in a neighbourhood) of the variable. This means a sharp peak in the distribution density function, see the graphs at the left hand of Fig. 7.

Therefore, the conditional probability density functions $p(x|\text{state } i)$ should be examined: does they have peaks? Unfortunately, in our septic shock data we can only observe the overall probability density function $p(x)$ which

has often a shape similar to a normal distribution. How do we know that it is composed by several normal distributions?

4.3 States and clusters

One of the main ideas for the construction of "states" is the observation that some observed variable values are more likely than the others. Typically, this is verified by a cluster analysis of the data. In a retrospective study Rixen et al. [6] observed that the k-means clustering of 17 cardiopulmonary and metabolic normalized variables of 1120 patient data sets (338 patients) resulted in 7 distinguishable states : (R) trauma recovery, (A) normal stress response, (B) metabolic insufficiency, (C1) early respiratory insufficiency, (C2) late respiratory insufficiency, (D) cardiogenetic decompensation, (H) hypovolemia without shock.

In difference to this, using the data of 656 intensive care unit patients (47 with a septic shock, 25 of them deceased) elaborated in a study made between November 1995 and December 1997 at the clinic of the J.W.Goethe-University, Frankfurt am Main [7], another study made in the same clinic between November 1993 and November 1995 and the data of the MEDAN [9] study were checked by a clustering analysis. The results were deceiving: no clusters were found except one huge cluster.

What can be concluded by this observation?

If several clusters exist, the existence of states would be supported, but since only one cluster is found, this does not tell anything about the existence of states. Why? If we cluster the set of all observations, we neglect the time structure of the observations. This might lead to an overlap of peak distributions giving rise to only one convex pdf hull of all variables.

If we include the time structure into the cluster analysis by using time as just another variable we do not solve the problem either. To see this, let us consider a system of three states. If we observe the time course of two patients, e.g. one going from state 1 to 3 and the other from 3 to 1, all pdfs are superseding, see the dotted line as hull of all pdfs in Fig. 8.

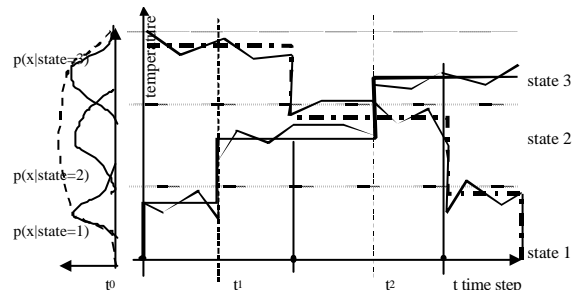


Fig. 8 Clustering the observations of a three-state-system

Although all observations are just random versions of an average state sign, the average of all state observations are not obliged to cluster at a given point of time. Thus, a broad interval of observations does not necessarily indicate the absence of states. Instead, we have to find other criteria to decide about the existence of states. This is up to future research.

Acknowledgement

We like thank especially Markus Pauen who prepared the simulation with the HTK tool for the patient data.

5. References

- [1] American College of Chest Physicians/ Society of Critical Care Medicine Consensus Conference (1992): Definitions for sepsis and organ failure and guidelines for use of innovative therapies in sepsis. *Crit. Care Med.* 20, 864-875.
- [2] Brause R., Hamker F., Paetz J., *Septic Shock Diagnosis by Neural Networks and Rule Based Systems*, In: L.C. Jain (ed.), *Computational Intelligence Techniques in Medical Diagnosis and Prognosis*, Springer Verlag, New York 2001, pp. 323-356
- [3] Bellman, R. (1961), *Adaptive Control Processes: A Guided Tour*, Princeton University Press.
- [4] Rabiner, Lawrence R.: A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proc. IEEE*, Vol. 77 (2), Feb. 1989, pp.257-286
- [5] Rangel-Fraustro M.S., Pittet D., Hwang T., Woolson R.F., Wenzel R.P.(1998): The dynamics of disease progression in sepsis: Markov modeling describing the natural history and the likely impact of effective antisepsis agents. *Clin. Infect. Dis.* 27(1), 185-190
- [6] Rixen D, Siegel J, Friedman H: Sepsis/SIRS, Physiologic Classification, Severity Stratification, Relation to Cytokine Elaboration and Outcome Prediction in Posttrauma Critical Illness, *J. Trauma: Injury, Infection, and Critical Care*, Vol 41, No.4, pp.581-598
- [7] Wade, S., Büssow, M, Hanisch, E. (1998), Epidemiology of Systemic Inflammatory Response Syndrome, Sepsis and Septic Shock in Surgical Intensive Care Patients, *Chirurg* 69, 648-655
- [8] The Hidden Markov Model Tool Kit, Cambridge University Engineering Department, <http://htk.eng.cam.ac.uk>
- [9] The MEDAN septic shock study, see <http://Medan.de>