Hidden Markov Models for Analyzing Medical Time Series in Order to Detect Nosocomial Pneumonia

UDC: 007:616-083 ; 519.863 ; 519.246.8

Marek Opuszko[1], Johannes Ruhland[1], Franziska Oroszi[2], Michael Hartmann[2] · Martin Specht[2], Friedrich-Schiller[1]
[1]*University Jena (FSU), Germany,Department of Information Systems*
marek.opuszko@uni-jena.de, j.ruhland@wiwi.uni-jena.de
[2]*University Hospital Jena, Germany*

XII International Symposium SymOrg 2010, 09 - 12 June 2010, Zlatibor, Serbia

Pneumonia - as an inflammatory illness of the lung - is a dangerous and often fatal disease. A special subclass, the ventilator associated pneumonia (VAP), is affecting up to one fifth of the patients at Intensive Care Units (ICU). Based on a two years dataset, collected at a large ICU, we investigate a new method for time series processing in order to develop an early warning system for developing pneumonia. The system focuses on the pre-onset phase of the disease to extrapolate the future's course. We utilized the functionality of Hidden Markov Models and the stacking paradigm to categorize and forecast given time series of a patient. Finally we demonstrate the benefits of our approach with a set of real patient data.

1. Introduction

The increasing use of information technology in hospitals and medical facilities of any kind led to an ongoing demand of exploiting the stored data to gain more knowledge about diseases, course of diseases, possible treatments methods etc. Especially critical environments like ICUs have a high sensibility for this issue. Here, 8 to 20% of the patients develop a VAP (N.N., 2005) what leads to mortality rates ranging from 20 to 50% or even 70% (N.N., 2005) (Heyland et al, 1999) (Tejerina et al, 2006). It is clear that an early and accurate diagnosis of VAP has a high relevance. An accurate diagnosis - and hence a faster recovery - shortens the patient's stay at the ICU and reduces both unnecessary stress for the patient and avoidable costs for the hospital (Oroszi, 2008). Besides this, physicians are overwhelmed by the massive amount of data recorded every day. Therefore new methods of utilization are needed. Data mining methods offer possibilities to transform pure and raw data into knowledge and can lead to a progress in treating methods.

This work is part of an interdisciplinary project between researchers from the department of information systems of Friedrich-Schiller University (FSU), Jena, an intensive care unit (ICU) and the hospital pharmacy of the same institution (Oroszi, 2008). The project, which followed the standard process model for data mining (CRISP-DM)(Chapman et al, 2000) has the aim to apply data mining techniques to the ICU database. In this paper, we will concentrate on time series processing, which is one important research scope of the overall project. The data we analyze are time series of an aggregated score value which have been generated during an early phase of the project (Oroszi, 2008). The particular goal of this work and other investigations within the project is to identify promising appendages for pneumonia prediction for further research. A crucial question regarding the project is: Are there differences between the pre-onset course of disease of patients with and without pneumonia? Furthermore the question arises, that if there are differences, are they trivial like "If the measured value reaches a certain point, the next day a pneumonia disease will be manifested" or if the course of disease contains more complex patterns to disclose? So if these patterns exist, data mining methods may be able to utilize them for an early warning system. A treating physician faces various types of data and information input to develop his diagnosis. With the increasing availability of computer generated and digitally stored data, treating physicians need tools to process this input in an effective and efficient way. An early warning system has to deliver reliable and intelligible information to support physicians in their daily work to ensure the best possible diagnosis through a diagnosis support system.

The paper is structured as follows:
In Section 2 we will give an overview of the given data and its structure. In Section 3 the theory and functionality of Hidden Markov Models is briefly illustrated. A simulation in Section 4 demonstrates the ability of the system to mirror a patient's stay at the ICU. Afterwards Section 5 illuminates our test arrangement to show how the components interact with each other. To complete the work we present the results of our investigations and highlight some perspectives for future research.

2. Data

All data was collected during an early phase of the project in the years 2004 and 2005 and was already preprocessed. The whole dataset exceeded more than 4,000 variables. Unfortunately there exists no single clinical manifestation to diagnose VAP, but several methods with diverging performance (Rea-Neto et al, 2008). We concentrated on time series of the clinical pulmonary infection score (CPIS) which have been calculated for every patient during the years 2004 and 2005. CPIS is a score value which was developed to ease the diagnosis of pneumonia and was first proposed by Pugin et al. in 1991. Although CPIS has some limitations regarding its moderate performance, it is a helpful tool in diagnosing VAP (Rea-Neto et al, 2008).

Table 1: *CPIS Input Valuation*

Input Feature	**Score Point**		
	0	**1**	**2**
Tracheal Secretions	Rare	Abundant	Purulent
Radiographic infiltrates	Absent	Patchy or diffuse	Localized
Fever (°C)	≥ 36.5 and ≤ 38.4	> 38.4 and ≤ 38.9	> 38.9 or < 36
Leukocytosis	≥ 4,000 and ≤ 11,000	< 4,000 or > 11,000	(> 4,000 or < 11,000) and ≥ 500 band forms
Oxygenation (PaO2/FIO2)	> 240 or accurate respiratory distress syndrome (ARDS)		≤ 240 and no ARDS
Microbiology	Negative		positive

As shown in Table 1 the CPIS is an integer score containing 6 score components (tracheal secretions, radiographic infiltrates, fever, leukocytosis, oxygenation, and semi-quantitive cultures of tracheal aspirates - microbiology) (Pugin et al, 1991). Every component adds an integer value between 0 and 2. Hence the CPIS has a maximum value of 12 - if all features add the value 2 - and a minimum of 0. According to international practice we consider pneumonia diagnosed, if the CPIS reaches the value ≥ 6 (Rea-Neto et al, 2008). The first day of pneumonia is in the following named "reaction day". Based on that convention, two groups of cases could be identified, cases with and without pneumonia. We will use this information later to evaluate our model. A first data overview showed a disadvantageous distribution especially in the group of cases with pneumonia. In this group we could extract 325 CPIS time series for the years 2004 and 2005. Due to the fact that the majority of the measured values occurred within a period after the reaction day was reached, only 79 time series with altogether 425 single CPIS values could be considered for processing in this group. According to our goal to analyze the pre-onset phase of pneumonia, data within this early period of the disease's course is essential. In the group of cases without pneumonia sufficient data was available. For instance, 147 cases with altogether 995 single CPIS values for the year 2004 and 138 cases with altogether 827 single CPIS values for the year 2005 could be extracted. Furthermore various time series contained gaps and were extremely sparse as many stays on the ICU were very short. Moreover the data was highly unbalanced as the group of non-pneumonic cases is extremely overrepresented. On top of that, the duration before a reaction day is reached was concentrated in the range of very short lengths in the group of pneumonic cases. Due to these limitations, there are methods required which can process this kind of data.

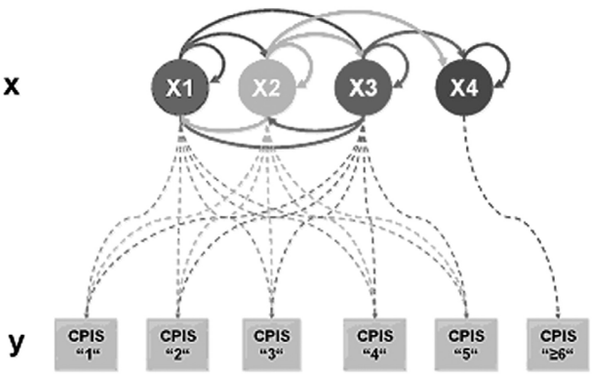
3. Hidden Markov models

The approach to investigate a possible forecast of pneumonia - based on CPIS time series - uses largely the functionality of the Hidden Markov Model (HMM). As many other methods have difficulties in processing time series of arbitrary and different length, HMM offer possibilities to process time series with a challenging characteristic. HMMs have been successfully used in data mining for many decades in speech recognition (Rabiner, 1989)(Manning and Schütze, 2005) as well as many other subjects, for instance bioinformatics (Gascuel and Moret, 2001) (Bystroff and Krogh, 2008). The mathematical description of HMM will largely follow Rabiner and Juang, 1986. The stochastic model of HMM is characterized by the combination of two random processes. An origin process with N different states $X = \{X1, .., XN\}$ is non-visible ("hidden"). This process can't be measured, but there exist M observable emissions $Y = \{y1, .., yM\}$ which offer in-

formation about the origin process. A patient's health state can be clearly understood as a random process. Within this process the patient's health state changes from time to time. Usually the health state is verbally described as "good", "bad" or "stable" etc. So obviously this process is hard to operationalize and hard to measure directly. Every treating physician uses symptoms and other available information to make a diagnosis about the patient's physical constitution. Now we make the assumption that the state of health can be represented by the hidden state of a HMM.

In a typical environment HMMs are used to classify temporal signals like continuous speech or gene sequences. Normally a signal is divided into blocks (frames) in a preprocessing step. In our case, one CPIS value represents one block. Due to the limitation that a CPIS measurement is only possible once a day, the frame rate is one day.[1] Hence a whole CPIS time series can be interpreted as emission symbols and represent the measurable symptoms. These time series allow estimations about the - hidden - origin process and furthermore the state of health. Picture 1 shows the structure of the HMM used to represent the course of disease. In this test arrangement, we assumed four origin states named: "green (X1)" for a stable state of health, "yellow (X2)" for an unstable state of health, "orange (X3)" for a dangerous condition of the patient and "red (X4)" for manifested. Three states (excluding "red") were considered the minimum necessary to model a disease development. The random variable x(t) is the hidden state at time $t(x(t) \in \{X1, X2, X3, X4\})$. Each state *X1,…, X4* has a transition distribution represented by the solid arks in Picture 1. The transition distributions of every state build the $N \times N$ transition matrix a_{ij} which is unknown until the model has been trained. Here a_{ij} is the probability of transitioning from state i to state *j* in the next step. Moreover the random variable *y(t)* is the emission at a *time* $t(y(t) \in \{y1, y2, y3, y4, y5, y6\})$. Every state has a probability distribution over the possible emissions *y1* to *y6*. The output probabilities - represented by the dotted lines in Picture 1 - build the $N \times M$ emission matrix , which defines the probability of every output token according to the actual hidden state of the model. So $b_i(k)$ is the probability of observing the token *yk* when the process is in state *i*. Furthermore a N-dimensional vector $\pi \in \{\pi 1, .., \pi N\}$ with initial probabilities for every state is given. Hence a HMM can be referred by λ , where $\lambda = (X, Y, a, b, \pi)$.

[1] **This is resultant to the temporal measure limitations of some components of the CPIS.**



Picture 1: *Structural graph of the HMM*

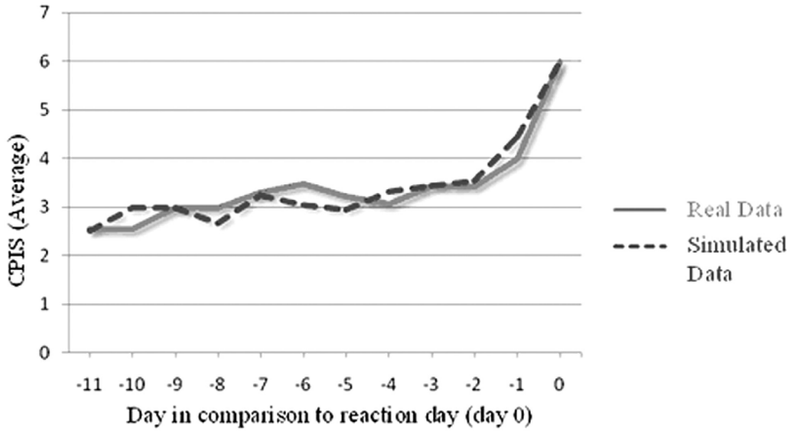
Regarding our assumptions in Section 2 we defined a CPIS of ≥ 6 equal condition "red". Furthermore we allowed no transitions from state "green" to state "red" and set all transition probabilities of state "red" to 0, except the transition to itself. The initial probability for state "red" was set to 0. Other assumptions or restrictions have not been made. According our research interest, four problems/issues regarding HMM arise:

1. Given a set of CPIS observations O with sequences $o_1,.., o_L$ and a HMM λ, how to adjust the model parameters *a, b* and π to maximize $P(O|\lambda)$ (We will from now on call this issue training). We will use the Baum-Welch algorithm (Baum et al, 1970) to solve this problem.
2. The most likely CPIS emission sequence, beginning at any point in time. We will use the forward-backward algorithm (Rabiner, 1989) to solve this problem and use this information to forecast the course of disease and predict the outbreak of pneumonia.
3. The most likely hidden state at a certain point in time and therefore also for a future emission. This problem is solved by using the viterbi algorithm.(Forney, 1973)(Rabiner, 1989) Thus we can gain more information about the course of a pneumonia disease.
4. The likelihood of a given CPIS sequence using the forward-backward algorithm which will be used to classify sequences. A description of this set of well known algorithms is given in Rabiner, 1989.

4. CPIS simulation

If proved that the stochastic properties of CPIS time series are represented correctly by the trained HMM, it would be possible to immediately start the forecast and count the correct predictions. To get an idea how accurate HMMs can model the development of pneumonia we assume that only if the model is able to simulate a

CPIS course, it can later predict that course properly. Therefore we trained a HMM as introduced in Picture 1 with a training set of cases with pneumonia. Picture 2 shows the average course of pneumonic patients and the average course of multiple simulations. The simulation for a particular sequence started with a first random emission and was stopped if a CPIS ≥ 6 was observed. In consideration of the fact that the generated sequences have different lengths, all sequences were aligned around the reaction day. Compared to the real time series the simulated sequences achieved a percentage deviation of 6.95% of average CPIS. This result makes clear, that the model is suited to adapt the stochastic properties of the given time series.



Picture 2: *Simulated vs. Real Time Series*

5. Test arrangement

The test arrangement shall adapt and simulate the course of pneumonia and deliver the likelihood of pneumonia in the future. If this functionality is achieved, the system could be transformed into an early warning or ”traffic light” system. According to that, the model could offer a decision support system for the treating physician to strengthen his diagnosis. The test arrangement is supposed to forecast a manifested pneumonia - condition ”red” - exactly one day before reaction day. For all other points in time the forecast shall be a state of ”green”, ”yellow” or ”orange”. The architecture of the model is guided by medical evidence that patients showing high pneumonia predisposition will contract and develop the disease much faster than average (Oroszi, 2008). The model mirrors this concept by implementing the stacking paradigm as shown in Picture 3. Two modules operate in series: one separating high- and low-risk patients (classification), the other doing the actual forecast for each of the two groups (prediction). The theoretical foundations of stacking have been outlined in relevant literature (Wolpert, 1992). Stacking is a method of using multiple serial or parallel models to achieve greater predictive accuracy (Ting and Witten, 1997). The design of the susceptibility prediction model was guided by the following requirements:

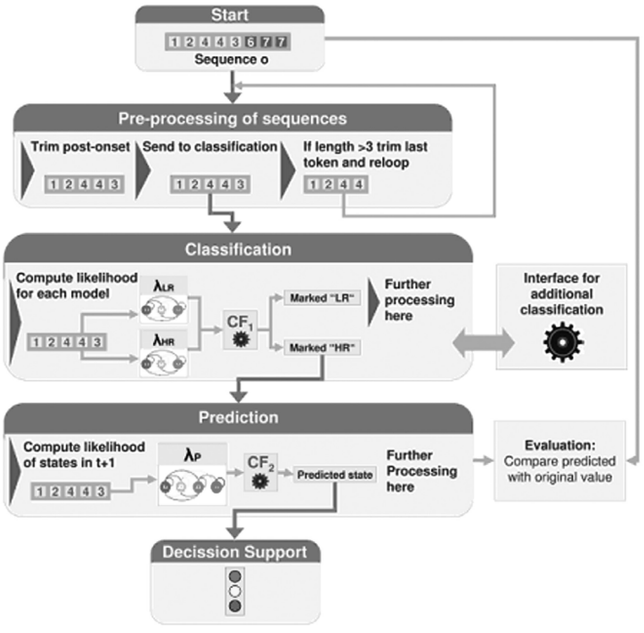
- Supervised learning should be used. This requires the concept of susceptibility to be broken down to observable quantities. As a first approximation, we used the ultimate outbreak of pneumonia according to the CPIS as a class variable. As an extension one might use a hidden variable that arises from Structural Equation Modeling (SEM) (see for example Buncher et al, 1991).
- The model should make probabilistic predictions with parameters that allow for easy adjustment of its α and β errors. It is the interaction in the different parts of the model that will eventually determine the system’s performance.
- The model must handle highly unbalanced learning samples, as the high susceptibility group is much smaller than the group of low susceptibility patients.

Classification The developed model is a combination of Hidden Markov Modeling and Bayesian reasoning. In the classification part, showed in Picture 3, two HMMs are derived, one for high risk patients (HR-patients - λ_{HR}) and one for the rest (LR-patients- λ_{LR}). In a Baum-Welch training each model adapted the characteristics of cases with pneumonia (λ_{HR}), respectively the cases without pneumonia (λ_{LR}). In both cases, the time series used for calibration will exclude the reaction day. Moreover the HMMs of the classification and the prediction model slightly differ in the number of states according to the fact that a condition ”red” does not exist in a pre-onset phase for the classification model. To classify a patient’s time series *o*, it will be fed into both calibrated HMMs, and the forward-backward algorithm will give the probability of this sequence occurring under the λ_{HR} and λ_{LR}-Model ($P(o|\lambda_{LR})$ and $P(o|\lambda_{HR})$ respectively). Bayes’ formula will yield $P(HR|o)$:

$$P(\lambda_{HR} \mid o) = \frac{P(o \mid \lambda_{HR}) \cdot P_{HR}}{P(o \mid \lambda_{HR}) \cdot P_{HR} + P(o \mid \lambda_{LR}) \cdot P_{LR}} \quad (1)$$

P_{LR} and P_{HR} are the usual a priori probabilities taken from general health statistics at the ICU. To get a valid result, the structure (*X, Y, N, M*) of both HMMs λ_{HR} and λ_{LR} has to be equal. Additionally a threshold level

or Certainty Factor CF_1 is introduced (see Picture 3). A *HR* classification thus is accepted, if the likelihood of ($P(o|\lambda_{LR})$ reaches a certain level. This offers the ability to set a minimum lower bound for accepting a *HR* classification. Focusing on *HR* cases in this work, the decision rule (which in turn will influence the α and β errors of this stage) is to label a case *HR*, if its a posteriori probability $P(HR|o)$ exceeds a threshold CF_1. Cases classified as *HR* will pass through the classifier. Strictly speaking, an analogue predictor/processor is needed for the *LR* cases as well. Due to substantially lower risk of developing pneumonia in this branch, we have not developed this model yet. At this point the benefits of the stacking paradigm become once more obvious, as every sequence classified correctly will directly decrease the number of β errors in the further steps. We return to the issue of unbalanced training data. Due to the fact that each HMM is trained independently, there is no restriction for equally balanced groups as long as sufficient training data exists. This illustrates once again the usability of HMM for classification.



Picture 3: *The test arrangement*

Prediction After a sequence o passed the classifier and was labeled "HR", a prediction model consisting of one HMM λ_P generates a forecast based on the characteristic of o. Having the four issues of Section 3 solved, it is easy to give a first one-day forecast from day t to day t + 1. From the time series of signals observed for each patient, the probability of the hidden state at t being i can be calculated. Knowing the (hidden) transition probabilities a_{ij} the probabilities for each hidden state at t + 1 can be calculated, which can be transformed into the probabilities for the observed emission at t + 1. Onset of pneumonia is forecasted, if the probability for reaching hidden state "red" or observed emission "CPIS > 6" exceeds a threshold, i.e.

– $P(\text{"red"}, t+1) > CF_2$ or
– $P(\text{"CP IS} > 6\text{"}, t+1) > CF_2$'

More sophisticated rules can be formed such as:
$P(\text{"red"}, t+1) > CF_2$"$P(x, t+1)$ or for all other hidden states x and a relative threshold of CF_2

We shall limit our discussion in this paper to the first case. The system could in principle be extended to give forecasts beyond $t + 1$. Given incubation times of two or three days for the case of pneumonia, this seems futile effort for the case of this disease, but is interesting from the general perspective. Summing up, we define a complex model $\kappa = \kappa(\lambda_C, \lambda_N, \lambda_P, CF_1, CF_2)$ containing all sub models and parameters. To finally measure the projection quality we split the whole dataset into a training set and a test set. The training set was used to train the classification models and the prediction model using the Baum-Welch algorithm. The system processed the test set and the prediction results could be compared with the real data[2] not known by the system. According to Picture 3 the whole process works as follows:

1. First of all, the system trims a test sequence o with the length T, leaving the pre-onset phase (o^* with length T^*). Time series with a pre-onset length ≤ 3 have been ignored due to a missing significance of short time series.
2. Clearly a decision support system shall not only predict the reaction day of pneumonia but also avoid false predictions before that day and in general for cases without pneumonia. A single pneumonia time series thus also provides snippets (o^*1, .., $o^*T^* - 2$) taken from times before the onset that should in turn be correctly identified as "no reaction day" series (reloop o^* to step 1).
3. Classification for "HR" and "LR". If sequence is tagged "HR", proceed with prediction. At this point additional methods and models, as mentioned before, may be plugged in.
4. Predict the next state according λ_P and CF_2.
5. Compare the predicted state with real state.

Tuning the system To finally run forecasts, the model faces another crucial question:
How shall Errors be treated correctly to gain an optimal result? The model has to predict a pneumonia at

[2] **Meaning the real condition the patient was in.**

the right time and avoid a false alarm before that point in time simultaneously, reflecting the α and β errors. Parameters CF_1 and CF_2 are the fundamental design parameters that may be tuned for "optimal" performance. A fundamental trade off may be seen between three types of errors:

- Error Class 1 - false negatives: The system has failed to identify the reaction day.
- Error Class 2 - false positives 1: The system has identified a reaction day for a pneunomic patient at the wrong point in time.
- Error Class 3 - false positives 2: The system has identified a reaction day for a non-pneunomic patient.

Medical practice will rank class 1-errors more serious than class 3 and both far more serious than class 2. In clinical terms, some class 3-errors might not even be regarded as ill-classified, as the distinction between pneunomia and other forms of pulmonary diseases such as severe bronchitis is blurred and the boundary defined by CPIS ≥ 6 is in reality a fuzzy one. In order to solve this problem we already introduced the two parameters (CF_1) and (CF_2). CF_1 and CF_2 can now be used to, for example, shift the arrangement to a more "false negative avoiding" behavior, paid with an increasing number of unnecessary treatments what is regarded less of a problem here. Now two possible ways in defining CF_1 and CF_2 exist, as a first solution, the treating physician may define the two parameters as a fixed constant. The disadvantage of this method is that a black-box like defining of the parameters is very abstract and not intuitive as the consequences are not obvious immediately. The other possibility is to optimize the parameters according to the given training data to get the best success-error relation. Therefore an optimal error-success relation has to be first operationalized. To solve this problem we introduce a target function $F = F(CF_1, CF_2)$ in which 3 different quality functions have to be weighted to meet the user's pretensions:

- Quality function QF_1, representing the percentage of correct predicted reaction days on conditions: model κ and values CF_1 and CF_2.
- Quality function QF_2, representing the percentage of false predicted reaction days for cases with pneumonia on conditions: κ, CF_1, CF_2.
- Quality function QF_3, representing the percentage of false predicted reaction days for cases without pneumonia on conditions: κ, CF_1, CF_2.

It is clear that all quality functions directly depend on the choice of CF_1 and CF_2. For instance, if CF_1 is set to 1, a classification to "HR" is virtually never accepted and therefore the reaction day will be hardly predicted. On the other hand no false prediction will occur. Additionally 3 parameters P_1, P_2 and P_3 are added to the target function to penalize the 3 quality functions if their values do not reach a minimum level. This can be used to set lower bounds for the arrangement's accuracy.[3]

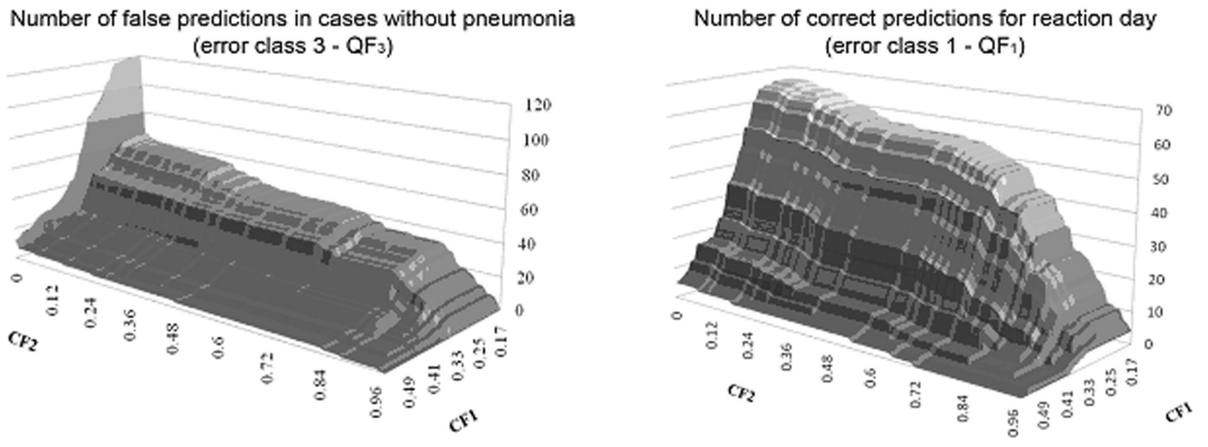
$$F(CF_1, CF_2) = w_1 \cdot QF_1(k \mid CF_1, CF_2) - w_2 \cdot QF_2(k \mid CF_1, CF_2) - w_3 \cdot QF_3(k \mid CF_1, CF_2) - P_1 - P_2 - P_3 \quad (2)$$

The target function F reflects the trade-off in having many predicted reaction days paid with many false predictions on the other hand. Now F provides an easy way to operationalize the user's needs in defining one's requirements. In a verbal expression a user may define: "The system has to identify at least 40% of all reaction days (P_1) and the correct prediction of reaction days (QF_1 and w_1) is twice as important as the avoidance of false predictions for non-pneumonic patients (QF_3 and w_3)". In this case, for instance, it makes sense to weight the avoidance of error class 1 (w_1 for QF_1) higher in order to prevent a missed treatment. With the definition of the weights w_1,.., w_3 the target function F delivers a definite target value for every combination of CF_1 and CF_2. Thus, the system can produce scalar optimal combinations of CF_1 and CF_2 according to the definitions of the weights. The definition of the weights for every parameter cannot be investigated in this paper as it is a complex medical decision. Instead of presenting one optimal solution we will demonstrate the error/success relation for any parameter setting and some exemplary results.

6. Experimental results

Picture 4 illuminates the effect of ? and ? error as it confronts the number of correct predicted reaction days and the number of falsely predicted reaction days (in the group of cases without pneumonia) depending on the value combination of CF_1 and CF_2. Obviously the shape shows a certain correlation which mirrors the α – β error trade off. With a decreasing number of correctly predicted reaction days, the number of false predicted reaction days decreases as well. Nonetheless, the shapes show differences which allow an adjustment of these two parameters. Especially the range of low values of CF_2 showed a dramatically increase in errors going along with a more or less stable number of correct predictions.

[3] **e.g. penalizing a result with to much failed predictions**



Picture 4: *Optimization surfaces of QF_3 and QF_1*

Table 2 depicts the results in every error class. In the first section we show how the system could decide whether a time-series belongs to a case with or without pneumonia based on different training methods. With a true positive rate around 82-83% the system is able to mark time-series as pneumonic or non-pneumonic. In the second section we used a set of 3 different weight settings to point out the mode of operation of our model in the α-β error trade off. As weight setting 3 (from an actual physician) exemplifies an increase of reaction day forecast, accuracy may be sacrificed in favor of a reduction of type 3 errors. As a reference we used two naive forecast strategies to compare the results. The first naive method predicts a reaction day in t + 1 if a CPIS value of ”5” is reached in t. The second naive method computes the slope of the time series based on the values of t and t – 1. As seen in Table 2 naive method 1 is in fact a rather powerful forecast that has strongly influenced the way the CPIS score is constructed. It is an advantage of our method that (through the choice of threshold CF_1 and CF_2) the value of α can be adjusted, which will also determine β (vice versa). The method moreover shows, that an adjustment of α errors is possible with simultaneously having a remarkable stable β error. The concrete trade off between the two types of errors may thus be explored.

Table 2: *Experimental results*

Classification model			
Training method		Cases with pneumonia (N=79)	Cases without pneumonia (N=285)
Standard Baum-Welch Training		82%	69.8%
Enhanced Training (Genetic Algorithm)		83%	73.5%
Overall results			
Method	Reaction Day correctly predicted	Error class 2: Reaction day was predicted too early	Error class 3: Reaction day way predicted wrongly
κ, Weight setting 1	41.6%	6.9%	6.5%
κ, Weight setting 2	63.8%	11.6%	9.3%
κ, Weight setting 3	43.2%	20.1%	6.5%
Trivial method 1 (y5, t → y6, t + 1)	44.3%	21.7%	13.8%
Trivial Method 2 (Slope)	31.6%	15.2%	10.4%

Summary

The test arrangement introduced in this work can be understood as a novel approach for processing and predicting medical data. However, the present research is a fist attempt to analyze pneumonia using HMM and has some limitations. The computations are based on a two years dataset with 79 (cases with pneumonia) and 285 (cases without pneumonia) time series only and should be tested with data from other years to evaluate the quality. Furthermore the data belongs to one ICU and it is not clear if the results could suffer from local effects and if other ICUs may show different results. In order to predict a VAP we concentrated on processing time series of cases with pneumonia. Reckoning the classification model, there is no subsequent processing for sequences marked with ”LR” at this point. In order to deploy a holistic system, further work at this point is needed. Furthermore the implications that leaded to the structure of the HMM (Number of states etc.) may be questioned and examined closer. The target function used to deliver an optimal result may be extended in order to consider economic issues like concrete cost rates for medication. Another benefit of this investigation is to corroborate knowledge on pneumonia disease. According to our results, the assumption of a short incubation time of pneumonia could be confirmed. On the other hand, the results reveal some limitations. Regarding the CPIS frame rate of one day, a prediction is strongly limited within this context. A higher frame rate would be helpful. Furthermore the system was based on an elaborate data preprocessing of two years data. Unfortunately this data is still quite insufficient due to the problems mentioned in Section 2. Hence, an integrated early warning system must be based upon a holistic a priori embedding in the hospital’s real time data infrastructure. If not already realized, the pre-use phase will demand a lot of resources. The installation of such a system is both time consuming and costly but warranted by its multiple uses. If such systems do exist, the pneumonia forecast itself drains very little resources in daily operations. It can and should be incorporated into a patient’s ”one paper” that gives an overview of the patient’s disease history which is a handy tool for treating physicians. Extensions in several points are evident. The system could be evaluated with other diseases with a higher incubation time. In the classification level other methods like Bayesian networks could be implemented. As other concepts of predisposition revealed high potential (Oroszi, 2008), further a priori methods should be explored, e.g. SEM. Within the framework of the stacked architecture of our system this can be easily achieved.

In conclusion, we have shown that data mining methods offer a high potential approach in disclosing a significant benefit in stored medical data. Apparently a fully automated ”out-of-the-box” solution will not deliver. Nonetheless, the system demonstrates how a stacked use of different methods enriches the disclosing of potentials, hidden in stored data.

BIBLIOGRAPHY

[1] Baum LE., Petrie T., Soules G., Weiss N. (1970) A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. The Annals of Mathematical Statistics 41(1):164–171, URL http://www.jstor.org/stable/2239727

[2] Buncher C.R., Succop P.A., Dietrich K.N. (1991) Structural equation modeling in environmental risk assessment. Environ Health Perspect 90:209–213

[3] Bystroff C., Krogh A. (2008) Hidden markov models for prediction of protein features. In: Protein Structure Prediction, Humana Press, Methods in Molecular Biology, vol 413, pp 173–198, DOI 10.1007/978-1-59745-574-9 7, URL http://www.springerlink.com/content/g4111p42750174r2/

[4] Chapman P., Clinton J., Kerber R., Khabaza T., Reinartz T., Shearer C., Wirth R. (2000) Crisp-dm 1.0 step-by-step data mining guide. Tech. rep., The CRISP-DM consortium, URL http://www.crisp-dm.org/CRISPWP-0800.pdf

[5] Forney G.D. (1973) The viterbi algorithm. Proceedings of the IEEE 61(3):268–278, URL http://ieeexplore.ieee.org/xpls/abs all.jsp?arnumber=1450960

[6] Gascuel O., Moret B.M.E. (eds) (2001) Algorithms in Bioinformatics, First International Workshop, WABI 2001, Aarhus, Denmark, August 28-31, 2001, Proceedings, Lecture Notes in Computer Science, vol 2149, Springer

[7] Heyland D.K., Cook D.J., Griffith L., Keenan S.P., Brun-Buisson C. (1999) The attributable morbidity and mortality of ventilator-associated pneumonia in the critically ill patient. The Canadian Critical Trials Group. Am J Respir Crit Care Med 159:1249–1256

[8] Manning C.D., Schütze H. (2005) Foundations of statistical natural language processing, 8th edn. MIT Press, Cambridge, Mass.

[9] NN (2005) Guidelines for the management of adults with hospital-acquired, ventilator-associated, and healthcare-associated pneumonia. Am J Respir Crit Care Med 171:388–416

[10] Oroszi F. (2008) Einsatz von Data Mining Verfahren auf medizinischen Daten - Anwendungsfall Pneumonie-Früherkennung. PhD thesis, Friedrich-Schiller-Universität Jena

[11] Pugin J., Auckenthaler R., Mili N., Janssens J.P., Lew P.D., Suter P.M. (1991) Diagnosis of ventilator-associated pneumonia by bacteriologic analysis of bronchoscopic and nonbronchoscopic "blind" bronchoalveolar lavage fluid. Am Rev Respir Dis 143:1121– 1129

[12] Rabiner L.R. (1989) A tutorial on hidden markov models and selected applications in speech recognition. In: Proceedings of the IEEE, pp 257–286

[13] Rabiner, L. R., and Juang, B. H. (1986), An Introduction to Hidden Markov Models, IEEE Acousfics, Speech & Signal Processing ,Magazine, 3, 1-16.

[14] Rea-Neto A., Youssef N.C., Tuche F., Brunkhorst F., Ranieri V.M., Reinhart K., Sakr Y. (2008) Diagnosis of ventilator-associated pneumonia: a systematic review of the literature. Crit Care 12:R56

[15] Tejerina E., Frutos-Vivar F., Restrepo M.I., Anzueto A., Abroug F., Palizas F., Gonzalez M., DÉmpaire G., Apezteguia C., Esteban A. (2006) Incidence, risk factors, and outcome of ventilator-associated pneumonia. J Crit Care 21:56–65

[16] Ting K.M., Witten I.H. (1997) Stacked generalization: when does it work. In: in Procs. International Joint Conference on Artificial Intelligence, Morgan Kaufmann, pp 866– 871

[17] Wolpert D. (1992) Stacked generalization. Neural Networks 5:241–259