Detecting Relationships Between Physiological Variables Using Graphical Modeling

Dr. Michael Imhoff,
Surgical Department, Community Hospital Dortmund, 44137 Dortmund, Germany
Dr. rer. nat. Roland Fried, Prof. Dr. rer. nat. Ursula Gather
Department of Statistics, University of Dortmund, 44221 Dortmund, Germany

Abstract

In intensive care physiological variables of the critically ill are measured and recorded in short time intervals. The proper extraction and interpretation of the information contained in this flood of information can hardly be done by experience alone. Intelligent alarm systems are needed to provide suitable bedside decision support. So far there is no commonly accepted standard for detecting the actual clinical state from the patient record. We use the statistical methodology of graphical models based on partial correlations for detecting time-varying relationships between physiological variables. Graphical models provide information on the relationships among physiological variables that is helpful e.g. for variable selection. Separate analyses for different pathophysiological states show that distinct clinical states are characterized by distinct partial correlation structures. Hence, this technique can provide new insights into physiological mechanisms.

Introduction

In intensive care detection of critical states and of intervention effects is of great importance for suitable bedside decision support. Clinical information systems can acquire and store physiological variables and device parameters online at least every minute. A physician can be confronted with more than 200 variables of each patient during his morning round[1] . In view of the high dimension of the data and the time critical situations physicians are steadily confronted with, severe problems arise from the natural limitations of human beings. It is difficult to develop a systematic response to problems involving more than seven variables[2], and human beings are not able to judge the degree of relatedness between more than two variables[3]. Thus, besides the aim of detecting clinical states, reducing the number of variables is a further task. Typically we select some of the variables using our personal experience. Of course, this is subjective, and it is important to know which and how much information we neglect in the reasoning process based on such a selection. Hence, we need information on the relationships between the variables. Graphical interaction models have become an important tool for investigating and modeling relationships within multivariate data. These models allow a simple and helpful graphical visualization, where the variables are represented by vertices and the relationships between the variables are illustrated by edges. Separations in the graph provide information on direct and indirect relationships in the data [4,5,6,7].

Methods

Data set. On the surgical intensive care unit of the Community Hospital Dortmund, a tertiary referral center, online monitoring data was acquired from 25 consecutive critically ill patients (9 female, 16 male, mean age 66 years) with extended hemodynamic monitoring requiring pulmonary artery catheterization, in one minute intervals with a standard clinical information system. These data were transferred into a secondary SQL database and exported into standard statistical software for further analysis. A total 129943 sets of observations were analyzed.

In the analysis we concentrated on the variables heart rate HR, arterial diastolic pressure APD, arterial systolic pressure APS, arterial mean pressure APM, pulmonary artery diastolic pressure PAPD, pulmonary artery systolic pressure PAPS, pulmonary artery mean pressure PAPM, central venous pressure CVP, blood temperature Temp and pulsoximetry SpO2. These variables are important for the detection of critical, possibly life-threatening situations as well as for intervention effects, and they provide information on the clinical status of the patient.

A first step for online monitoring is to find a manageable number of representative variables. In clinical practice, we typically select some of the observed variables and base our decisions on them. However, to get reliable and interpretable results without substantial loss of information we need to understand the relationships between the variables. In order to explore the use of graphical models for this task, we analyzed the relationships between all vital signs mentioned above. Since we want to get a general impression we estimated the partial spectral coherences from the full data set for each patient.

We make use of a given classification of observation periods for each patient into different clinical states to evaluate whether these states can be characterized by different partial correlation structures.

Graphical interaction models. Between multiple variables usually a multitude of relationships exists, but many of them are indirect, i.e. they are induced by

others. Distinguishing between direct and induced relationships among the observed variables and primary and secondary consequences of medical interventions is difficult from experience alone. Statistical analysis in form of graphical models helps to reveal the essential relationships which are not induced by others. Visualization of a graphical model is accomplished by a graph: We draw a circle for each variable and connect each pair of variables by an edge whenever the relation between these variables persists after conditioning on all the other variables. In this way the indirect character of some marginal relationships which are induced by underlying conditional dependencies can be illustrated. Indirect relationships can result from successively ordered direct influences. We concentrate on undirected graphs where undirected edges (simple lines) represent symmetrical interactions between the variables. A medical example for the use of directed graphs where directed edges (arrows) represent directed influences can be found in [8].

In the following we compare "empirical relationships" found by statistical analysis to "physiological relationships" based on medical knowledge. Physiological relationships mean that a change in one physiological variable leads to a change in another physiological variable. For instance, systolic, diastolic and mean blood pressures are different representations of the same physiological process, i.e. a pulsatile blood pressure. Therefore, an increase of systolic blood pressure close to always leads to an increase in mean blood pressure. This is something that can directly be deducted from physiological knowledge. In other situations the direction of the relation may change in different disease states. For instance, during volume depletion an increase of CVP and PAPM will typically lead to an increase in APM and a decrease in HR, while in congestive heart failure under high doses of inotropes an increase in HR may typically lead to an increase in APM. Hence the term physiological relationship does not imply any causal, linear or non-linear relation, nor a direction for it.

From a statistical point of view, measurements of physiological variables observed in short time intervals form multivariate time series as there may be interactions not only between the measurements observed at the same time point, but at short time lags, too. Therefore we use partial correlation graphs for multivariate time series[9,10,11]. Here, linear relationships between every pair of variables at all time lags are investigated controlling for the linear effects of the other variables at all time lags, i.e. after all linear effects of the other series have been removed[12]. These relationships are called partial correlations and can be expressed equivalently in the frequency domain using the partial spectral coherence, that measures the partial correlations at all frequencies. Hence, partial correlation graphs allow to detect relationships in form of partial linear, possibly time-lagged dependencies between the variables of a multivariate time series. Moreover, under some weak regularity assumptions we can interpret the separations found in a graphical model.

For estimation of the partial spectral coherences from multivariate time series representing physiological variables we use the program "Spectrum" developed by Dahlhaus and Eichler[13]. Based on the estimated partial spectral coherence a decision has to be made on whether the true partial spectral coherence may be identical to zero because sampling variability always causes estimates to be distinct from zero. Simple testing results in a crude "yes-no" statement, while it is well-known that physiological relationships may differ in strength. Hence, we decide to measure the importance of the edges in a simple way. We decide to classify the relationships into high, medium, low and zero partial correlation on the basis of the area under the estimated partial spectral coherence. Between 366 and 10929 observations could be used for the estimation of the partial spectral coherences for each patient. Using the clinical data we now test whether graphical interaction models can reliably identify known strong relationships, e.g. between systolic, mean and diastolic blood pressures, and known likely relationships, e.g. between HR, PAP, and CVP.

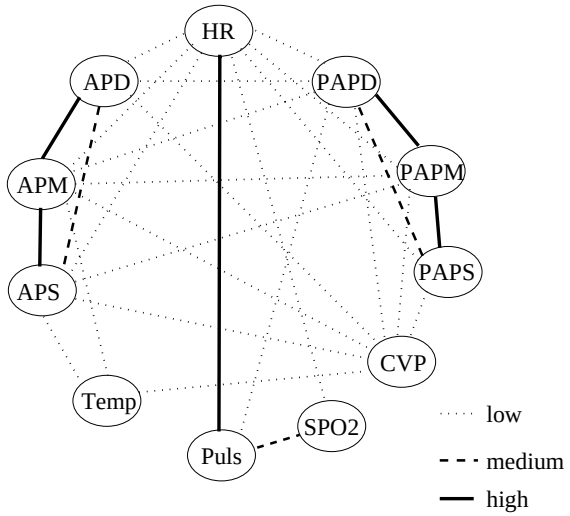
Results

As expected, for all patients strong relationships could be identified between the arterial pressures (APS, APD and APM), between the pulmonary artery pressures (PAPS, PAPD, PAPM) as well as between heart rate and pulse, the strength of the relation between the systolic and the diastolic pressure being always smaller than between each of these and the corresponding mean pressure. These relationships measured by the area below the partial spectral coherences were found to be much stronger than all other relationships. Hence, we can identify groups of strongly related variables from an analysis of the full data sets. Further relationships could be identified for some patients, e.g. between arterial pressures and heart rate, and between pulmonary artery pressures and central venous pressure. Figure 1 shows a typical example of a partial correlation graph for the hemodynamic system resulting from the analysis of the data measured for one patient.

A partitioning of the variables into strongly related subgroups as given above can be used to reduce the number of variables which have to be considered for online monitoring, i.e. for variable selection. The absence of a relation between two variables V1 and V2 means that the observations of V2 do not add anything to explain the course of variable V1 (and vice versa) given the measurements of the remaining vari-

ables. On the other hand, if a variable has strong relationships to several other variables it provides a lot of information on these variables. Selecting APM from the strongly related subgroup of arterial pressures and neglecting APD and APSYS for clinical monitoring is therefore meaningful from a statistical point of view. The same applies to pulmonary artery pressures.

Figure 1: Partial correlation graph. Different line types depict different strength of correlation.



In the previous analysis we inspected the relationships using all variables. This may hide some relationships when there are groups of variables which are only slightly different representations of the same physiological process. For instance, APM is a transform of APS and APD, and PAPM is a transform of PAPS and PAPD. When analyzing whether there is a relationship between PAPM and APM, we subtract the linear influences of all other variables including APD, APS, PAPD and PAPS. After elimination of the linear effects of these variables, the remaining variability for both APM and PAPM is very low. This possibly masks some weaker relationships. In consequence, systolic and diastolic pressures are dropped in the following, which is in line with medical reasoning as mentioned above. In this way a set of 'important variables' consisting of HR, APM, PAPM, CVP, SPO2 and Temp is retained.

For these variables all observations without missing values are included in the analysis for each patient. Table 1 summarizes the results. It should be noted that the classifications of the relationships are not independent for the same patient. The strongest partial correlations are observed between pulmonary artery and central venous pressures. Strong partial correlations can also be detected between these intrathoracic pressures and mean arterial pressure, and between heart rate and mean arterial and pulmonary arterial pressures. The weakest partial coherences are between body temperature and all other vital signs.

Table 1: Summary of the classifications based on the partial spectral coherences (25 patients).

	High	Medium	Low	Zero
HR-APM	3	3	13	6
HR-PAPM	1	3	11	10
HR-CVP	0	2	13	10
HR-SPO2	1	1	7	16
HR-Temp	0	0	14	11
APM-PAPM	2	10	13	0
APM-CVP	0	0	17	8
APM-SPO2	0	3	12	10
APM-Temp	0	1	12	12
PAPM-CVP	14	9	2	0
PAPM-SpO2	0	1	7	17
PAPM-Temp	0	1	14	10
CVP-SpO2	0	2	11	12
CVP-Temp	0	3	7	15
SpO2-Temp	0	0	8	17

Distinct clinical states such as pulmonary hypertension, septic shock, congestive heart failure and vasopressor support are accompanied by different pathophysiological responses of the circulatory system. These changes may be supposed to result in differences in the interactions between the vital signs, too. For instance, pulmonary hypertension can be characterized by an elevated PAPM. Due to the increased right ventricular pressures we expect strong interactions between CVP and PAPM. On the other hand, the increased pulmonary vascular resistance may attenuate the interactions between PAPM and APM as changes in PAPM will have a less than normal effect on left ventricular preload. In consequence, the expected associations of vital signs show a different picture for the state of pulmonary hypertension than the relationships under normal physiological conditions.

In the following we investigate, whether graphical correlation models can detect differences in the status of the circulatory system. Therefore, for each patient the predominant pathophysiological state is determined for every time point from the medical record. Then the time series are partitioned into segments of at least 300 observations during which the patient is in the same clinical state. The partial spectral coherences between the vital signs are estimated separately for each of these segments. Tables 2 to 5 summarize the results for the distinct clinical states.

Although the number of samples is small, there are obvious differences between the partial correlation patterns for distinct clinical states. Most of these dif-

ferences can be explained by known physiological mechanisms. While for most states strong partial correlations can be found between heart rate and blood pressures as well as between intrathoracic and arterial pressures, the status of pulmonary hypertension is predominantly associated with strong partial correlations between the intrathoracic pressures only.

Table 2: Summary of the classifications for congestive heart failure (15 segments).

	High	Medium	Low	Zero
HR-APM	0	5	8	2
HR-PAPM	0	2	7	6
HR-CVP	0	1	4	10
HR-SPO2	0	0	8	7
HR-Temp	0	2	6	7
APM-PAPM	3	11	1	0
APM-CVP	0	4	5	6
APM-SPO2	0	0	10	5
APM-Temp	0	1	6	8
PAPM-CVP	4	10	1	0
PAPM-SpO2	0	1	7	7
PAPM-Temp	0	1	5	9
CVP-SpO2	0	2	3	10
CVP-Temp	0	0	7	8
SpO2-Temp	0	3	6	6

Table 3: Summary of the classifications for pulmonary hypertension (17 segments).

	High	Medium	Low	Zero
HR-APM	0	7	8	2
HR-PAPM	0	1	11	5
HR-CVP	0	2	9	6
HR-SPO2	0	0	5	12
HR-Temp	0	0	9	8
APM-PAPM	0	7	9	1
APM-CVP	0	0	10	7
APM-SPO2	0	2	8	7
APM-Temp	0	1	9	7
PAPM-CVP	7	9	1	0
PAPM-SpO2	0	0	6	11
PAPM-Temp	0	1	8	8
CVP-SpO2	0	1	11	5
CVP-Temp	0	1	7	9
SpO2-Temp	0	0	5	12

Table 4: Summary of the classifications for septic shock (19 segments).

	High	Medium	Low	Zero
HR-APM	3	6	6	4
HR-PAPM	0	1	11	7
HR-CVP	0	2	10	7
HR-SPO2	0	3	8	8
HR-Temp	0	2	9	8
APM-PAPM	0	8	7	4
APM-CVP	0	2	10	7
APM-SPO2	0	1	10	8
APM-Temp	0	0	12	7
PAPM-CVP	9	7	3	0
PAPM-SpO2	0	2	10	7
PAPM-Temp	0	2	10	7
CVP-SpO2	0	2	9	8
CVP-Temp	0	2	11	6
SpO2-Temp	0	2	13	4

Table 5: Summary of the classifications for vasopressor support (6 segments).

	High	Medium	Low	Zero
HR-APM	3	3	0	0
HR-PAPM	1	1	2	2
HR-CVP	1	3	2	0
HR-SPO2	1	3	2	0
HR-Temp	0	0	4	2
APM-PAPM	4	1	1	0
APM-CVP	2	1	2	1
APM-SPO2	0	4	2	0
APM-Temp	0	0	3	3
PAPM-CVP	2	2	1	1
PAPM-SpO2	0	1	2	3
PAPM-Temp	0	2	2	2
CVP-SpO2	0	2	2	2
CVP-Temp	0	0	2	4
SpO2-Temp	0	1	2	3

For the clinical status of congestive heart failure we find high partial correlations between APM and PAPM. This can be explained by a failure of the left ventricle, where the forward failure of the left ventricle leads to a decrease in APM, and the concurrent backward failure to an increase in PAPM. For the status of vasopressor support there are strong partial

correlations between APM and PAPM, too. But in comparison to the previous states there are also higher partial correlations between HR and APM. This is due to the therapy which inhibits the normal autoregulation of the circulatory system. Hence, there are strong positive interactions between APM and PAPM, while the influence of CVP on the other variables is reduced.

Conclusion

We found that graphical partial correlation analysis based on partial spectral coherences can reliably detect known physiological relationships between hemodynamic variables. The insights gained by this method are useful to improve online monitoring of vital signs since it allows an improved application of methods for dimension reduction. One possibility is to select suitable subsets of important variables from the graphs. Alternatively, we can deduce information on the partial correlation structure from the partial correlation graph to enhance methods such as principal component analysis or factor analysis for time series, which are also accomplished via spectral analysis[12]. Multi-block principal component analysis has been suggested to monitor high-dimensional time series[14]. Here, the variables are organized in subsets to which principal component analysis is applied separately. Such subsets of closely related variables can also be identfied from graphical models for historic data. In our study the blocks obtained from the data analysis agree with the blocks obtained from medical knowledge. Therefore, we expect to gain reliable insights when applying this methodology to time series describing other variables, for which we have less medical knowledge so far. Special partial correlation patterns may even represent specific pathophysiological states in the critically ill. This can be useful for reaching deeper insights into the causes of clinical complications as well as for detecting such complications by additional data analysis. In view of the high sampling rates of modern equipment this method could even be applied to detect these complications online in the time critical situations on the intensive care unit. Further studies are projected to validate these preliminary findings with larger groups of patients.

In summary, graphical partial correlation analysis seems to support the analysis of correlations in multivariate physiological time series. As the final statistical analysis results in simple graphs, interpretation of the partial correlation patterns can be accomplished by physicians without further statistical knowledge.

References

1. Morris G, Gardner R Computer applications. In: Hall J, Schmidt, G, Wood, L editors. Principles of critical care. New York: McGraw-Hill 1992, pp. 500-514.
2. Miller G. The marginal number seven, plus or minus two: Some limitations to our capacity for processing information. Psychol Rev 1956; 63: 81-97.
3. Jennings D, Amabile T, Ross L. Informal assessments: Data-based versus theory-based judgments. In: Kahnemann D, Slovic P, Tversky A. editors. Judgments Under Uncertainty: Heuristics and Biases. Cambridge: Cambridge University Press 1982; pp. 211-230.
4. Whittaker J. Graphical Models in Applied Multivariate Statistics. Wiley: Chichester, 1990.
5. Cox DR, Wermuth N. Multivariate Dependencies. Chapman & Hall: London, 1996.
6. Lauritzen SL. Graphical Models. Clarendon Press: Oxford, 1996.
7. Edwards D. Introduction to graphical modelling. Second edition. Springer: New York, 2000.
8. Tritchler D. Reasoning about data with directed Graphs. Statistics in Medicine 1999; 18: 2067-2076.
9. Brillinger DR. Remarks Concerning Graphical Models For Time Series And Point Processes. Revista de Econometria 1996; 16:1-23.
10. Dahlhaus R. Graphical interaction models for multivariate time series. Metrika, 2000.
11. Gather U, Imhoff M, Fried R. Graphical models for multivariate time series from intensive care monitoring. Statistics in Medicine, to appear.
12. Brillinger DR. Time Series. Data Analysis and Theory. San Francisco: Holden Day 1981.
13. Dahlhaus R, Eichler M. Spectrum. The program is available at http://www.statlab.uni-heidelberg.de/projects/graphical.models/
14. MacGregor JF, Jaeckle C, Kiparissides C, Koutoudi, M. Process Monitoring and Diagnosis by Multiblock PLS methods. AIChE J. 1994; 40: 826-838.

Supported, in part, by the DFG (SFB475)