

# Data Mining auf Intensivdaten zur Modellierung von Intensivpatienten am Anästhesie-Simulator

Jochen Vollmer<sup>1</sup>, Stefan Mönk<sup>1</sup>, Wolfgang Heinrichs<sup>1</sup>, Thomas Uthmann<sup>2</sup>

<sup>1</sup>Klinik für Anästhesiologie, Klinikum der Johannes Gutenberg-Universität, Mainz  
55131 Mainz

E-Mail: vollmer@anaesthesie.klinik.uni-mainz.de

<sup>2</sup>Institut für Informatik, Johannes Gutenberg-Universität, Mainz  
55131 Mainz

## Abstract

Zur Vorbereitung der Modellierung verschiedener Typen von Intensivpatienten am Patientensimulator wurden mittels Clustering- und Assoziationsanalyse-Methoden eine Population von Intensivpatienten hinsichtlich ähnlicher Patienten und typischer Verläufe von Krankheiten untersucht.

Es wurde eine Zerlegung der Patientenpopulation in 6 medizinisch relevante und interpretierbare Klassen gefunden. Die Analyse von Verlaufsdaten ergibt in einem ersten Ansatz einfache medizinische Aussagen, die Anlass zur weiteren Untersuchung geben.

## 1. Einleitung und Motivation

Seit 1997 betreibt die Klinik für Anästhesiologie der Joh. Gutenberg-Universität, Mainz, ein Simulationszentrum zur Aus- und Weiterbildung von Anästhesisten. Simulationstraining als Ausbildungswerkzeug in der Medizin stellt ein relativ neues Konzept dar. Parallelen zwischen Konzepten des Simulationstrainings in der Luftfahrt und der Problematik in der Anästhesie lassen sich relativ leicht ziehen und Konzepte zur Bewältigung von Krisensituationen übertragen<sup>i</sup>.

Möglich wurden Simulationskurse in der Anästhesie jedoch erst durch Entwicklung von sogenannten Fullscale Simulatoren, die im Gegensatz zu schon lange eingesetzten Tasktrainern, die nur einen eng umgrenzten Bereich einer medizinischen Problemstellung darstellten, das komplexe Zusammenspiel verschiedener Organgruppen zu einer Art Gesamtorganismus an einer menschenähnlich gestalteten Puppe darstellten. Eine wichtige Eigenschaft solcher FullScale Simulatoren liegt in der Steuerung der Simulatoren. Sie verfügen über mathematisch physiologische Modelle, wie ein Herz-Kreislauf- und Lungenmodell sowie ein pharmakologisches Modell, die den Simulator auf eine Intervention eines Teilnehmers einer Simulationsschulung reagieren lässt.<sup>ii</sup>

Technisch betrachtet besteht der in Mainz benutzte Simulator<sup>iii</sup> aus einer Computersimulation, die über das Zusammenspiel von Hardwareeinheiten, die errechneten Parameter der Simulationsmodelle an einer Patientenpuppe darstellt. Eine realistische Arbeitsumgebung wird durch die Benutzung kliniküblicher Beatmungsgeräte und Überwachungseinheiten sowie die Installation des Simulators in der Umgebung eines echten Operationssaals erreicht.

Die Reaktionsweisen eines solchen Simulators sind jedoch nur so gut, wie die physiologischen Modelle, die im Hintergrund der Simulation ablaufen. Die Physiologie „gesunder“ Patienten, d.h. Patienten, die außer einer akuten Erkrankung über einen gesunden Körper verfügen, ist hinreichend mathematisch modelliert und kann somit in einer

Computersimulation eingebaut und verwendet werden. Dies reicht aus, um intraoperative Problemsituationen gut darzustellen.

Komplexer gestaltet sich jedoch die Simulation von schwer und schwerst kranken Patienten, wie sie neben den postoperativen Überwachungspatienten auf Intensivstationen anzutreffen sind. Das Gebiet der Intensivmedizin stellt jedoch für den Anästhesisten ein wichtiges Betätigungsfeld dar.

Simulationsschulungsangebote, die den Umgang und die Therapie dieser Patienten behandeln, wären von großem Interesse und werden bisher nicht angeboten. Dies liegt daran, daß Patienten mit schweren Organschädigungen oder kritischem Gesamtzustand oft von der gut modellierten Physiologie im Prinzip gesunder Patienten abweichende Reaktionen aufweisen und wenige mathematische Modelle zur Beschreibung der Pathophysiologie als Grundlage der Simulation bekannt sind<sup>iv</sup>.

Weiterhin spielen Langzeitregelmechanismen wie Hormonschwankungen oder Blutdruckveränderungen im Verlauf eines Tages für die Darstellung eines Patienten in einer Operationsumgebung keine wesentliche Rolle und sind deshalb oft in der Modellation an Simulatoren nicht berücksichtigt. Diese Effekte gewinnen aber für die Beschreibung des Zustandes eines Intensivpatienten eine starke Bedeutung.

Zudem sind bei Intensivpatienten von anderen Patienten abweichende Reaktionen auf Medikamentengaben, die eventuell auf einen Gewöhnungseffekt hinweisen können, zu beobachten und Patienten mit oft verschiedener Grunderkrankung weisen einen ähnlichen Krankheitsverlauf auf.

Trotz der Vielfalt der Verläufe, Krankheitsbilder und der fehlenden mathematischen Modellierung dieser Patienten, können erfahrene Ärzte oft den Zustand und die Reaktionsweisen ihrer Patienten einschätzen und errahnen.

Aus Sicht der KI lässt sich solches Wissen beispielsweise in der Modellierung eines Expertensystems abbilden. Ein anderer Ansatz stellt die Analyse von relevanten Daten durch Methoden des Data Minings dar. Im weiteren soll dieser Ansatz verfolgt werden.

Lassen sich Erfahrungswerte und Einschätzungen eines Arztes an Parameterkombinationen über Daten der betrachteten Patienten festmachen? Welche Arten von Daten sind hierfür nützlich?

## **2. Daten von Intensivstationen:**

Schwerkranke Patienten auf Intensivstationen werden durch eine Vielzahl von Geräten und Personal überwacht und eine Flut von Parametern, die den Zustand des Patienten beschreiben, fallen an.

Seit den 90er Jahren nehmen sogenannte Patienten-Daten-Management-Systeme (PDMS) verstärkt Einzug auf Intensivstationen, die auf der einen Seite organisatorische Daten der Patienten aufnehmen, auf der anderen Seite klinische Parameter über den Zustand der betreuten Patienten direkt aus angeschlossenen Medikamentenpumpen, Monitorgeräten oder Beatmungsgeräten erfassen.

Die Arbeit mit diesen Daten birgt mehrere Problemquellen: Nicht jeder Patient ist gleich überwacht. Damit fallen verschiedene Parametermengen für Patienten an und die Vergleichbarkeit der Patienten wird dadurch beeinträchtigt. Zum anderen werden Rohdaten aufgezeichnet, die eventuell außerhalb des Kontextes der Eigenart des Patienten nicht verständlich sind. So lassen sich beispielsweise Artefaktsituationen aus solchen Rohdaten nicht nachvollziehen. Weiterhin weisen die Daten von PDMS oft eine schlechte Datenqualität auf. Beispielsweise erzeugt der Einsatz von älteren Geräten oder Geräten, deren Parameter nicht automatisiert durch das PDMS aufgezeichnet werden können, unvollständige Datenreihen.

Eine andere Art der Datenerfassung wurde mit dem hier im weiteren benutzten Datenstamm verfolgt:

Benutzt wurde eine Datenbank aus 2300 Patienten, die an 15.000 Behandlungstagen erhobene Daten bundesweit mehrerer beteiligter Intensivstationen von 1996 umfasst. Die Daten wurden ursprünglich im Zusammenhang qualitätssichernder Untersuchungen<sup>Y</sup> erhoben und umfassen Daten, die zum einen den Aufwand der Behandlung, Datenfelder zu Personen-bezogenen Daten des Patienten, verschiedene a priori Risikoklassifizierungen Aufnahme- und Entlassungsdiagnosen, Felder zu Vorerkrankungen und Vorbehandlungen sowie den Entlassungszustand des Patienten beschreiben. Zum anderen werden Parameter erfasst, die den zeitlichen Krankheitsverlauf des Patienten charakterisieren.

Bei diesen Parametern handelt es sich um alle 24 Stunden von einem Arzt erhobene Daten, die den Zustand von 14 betrachteten Organsystemen durch den Grad der Funktionseinschränkung des betrachteten Organs wiedergeben. Somit handelt es sich hierbei um von einem Fachmann vorinterpretierte Daten, die den beschriebenen Nachteil von Rohdaten nicht aufweisen. Die Zahl der Parameter ist im Vergleich mit Daten eines PDMS als recht gering und die Datenqualität durch in das Erfassungswerkzeug integrierte Plausibilitätsprüfungen als relativ hoch anzusehen.

### 3. Fragestellung und Methoden

Im Bezug auf die Vorbereitung der Simulation von Intensivpatienten ist die Suche nach charakteristischen Gruppen von Patienten auf Intensivstationen und Aufschlüsse über typische Verläufe von Krankheitsbildern von Interesse. Diese Erkenntnisse sollen helfen Patientenprofile und Szenarien zur Simulation zu entwerfen.

Zur Trennung der für diese Betrachtungen uninteressanten postoperativen Überwachungspatienten von den schwierig zu modellierenden beschriebenen Patienten wurde die Patientenpopulation nach dem Kriterium der Liegedauer der Patienten unterteilt. Eine Liegedauer von mehr als 5 Tagen wurde nach Einschätzung des Experten und Analyse der Verteilung der Liegedauer als Trennkriterium für sinnvoll betrachtet. Damit reduzierte sich die Anzahl der betrachteten Patienten auf 446. Zur Bearbeitung der Fragestellungen wurden dann die Daten in zeitabhängige und zeitunabhängige Parameter aufgespalten und getrennt betrachtet. Auf den zeitunabhängigen wurden mittels der Methode der Clusteranalyse Zerlegungen der Patientenpopulation untersucht und auf den zeitabhängigen mit Hilfe von Assoziations- bzw. Sequenzanalysen nach Aufschlüssen über den zeitlichen Verlauf der Organfunktionen über den Behandlungszeitraum gesucht.

Als Werkzeug für diese Datenanalysen kam der SAS Enterprise Miner V3<sup>vi</sup> zum Einsatz.

### 4. Resultate:

Mit Hilfe der Cluster-Methode konnte eine Zerlegung der Patientenpopulation in 6 Cluster über den Aufnahmedaten und –diagnosen erreicht werden. Diese Auswahl der Variablen schien aus zwei Gründen sinnvoll. Zum einen stellen diese genau die Parameter dar, die einem Arzt bei der Aufnahme und der damit ersten Einschätzung eines Patienten zur Verfügung stehen, zum anderen geben diese Parameter ein verständliches Schema zur Modellierung von Simulationspatienten wieder.

Die Cluster stellen nach klinischen Gesichtspunkten Gruppen von Patienten dar, denen in der Realität existierende Krankbilder zugeordnet werden können. Zur Veranschaulichung wird ein Krankheitsbild als Assoziationshilfe zur Beschreibung der einzelnen Cluster hinzugefügt.

Bei der Zerlegung der Patientenpopulation werden nicht zu kleine, d.h. zu spezielle Gruppen gefunden. Seltenerer Fälle von Krankheiten werden somit allgemeineren Gruppen von Krankheitsbildern zugeordnet und können aus diesen Clustern nicht erkannt werden. Die

Clustergröße variiert zwischen minimal 36 zugeordneten Patienten von Cluster 2 und maximalen 110 Patienten von Cluster 1.

Die Reihenfolge der in der Tabelle aufgeführten charakterisierenden Kriterien versucht die beim Clustering angegebene Trennschärfe der Variable auf die Clusterbildung wiederzugeben.

Abnehmende Trennschärfe ↓	<b>Cluster 1</b>	<b>Cluster 2</b>	<b>Cluster 3</b>
	<ul style="list-style-type: none"> <li>- Ungeplante Aufnahme</li> <li>- Keine Operation</li> <li>- Schädigungen an der Lunge</li> <li>- Schädigungen am Kopf</li> <li>- Niedrige ASA-Klassifikation</li> <li>- Wenig Vorbehandlung</li> <li>- Einlieferung von außen.</li> </ul>	<ul style="list-style-type: none"> <li>- Schädigungen an Extremitäten</li> <li>- Einstufung als aufwendiger Patient</li> <li>- Keine Schädigung am Herz</li> <li>- Keine Schädigung am Kopf</li> <li>- Kein Stoffwechselproblem</li> <li>- Patient ist eher jung</li> </ul>	<ul style="list-style-type: none"> <li>- Älterer Patient</li> <li>- Weiblicher Patient</li> <li>- Stoffwechselprobleme</li> <li>- Erkrankung in Bauchregion</li> <li>- Keine Schädigung der Lunge</li> <li>- Hohe ASA-Klassifizierung</li> <li>- Verlegung innerhalb der eigenen Klinik</li> </ul>
	Bsp.: Patient mit akutem Lungenversagen	Bsp.: Traumapatient	Bsp.: Patientin mit Darmverschluss
Abnehmende Trennschärfe ↓	<b>Cluster 4</b>	<b>Cluster 5</b>	<b>Cluster 6</b>
	<ul style="list-style-type: none"> <li>- Hohe ASA-Klassifizierung</li> <li>- Postoperative Einlieferung</li> <li>- Kardiales Problem</li> <li>- Aufnahme von Extern</li> <li>- Alter nicht charakteristisch</li> </ul>	<ul style="list-style-type: none"> <li>- Ungeplante Aufnahme</li> <li>- Diagnosen betreffen den Bauchbereich</li> <li>- Diagnosen betreffen die Lunge</li> <li>- Keine Schädigung am Kopf</li> <li>- Alter nicht charakteristisch</li> <li>- ASA nicht charakteristisch</li> </ul>	<ul style="list-style-type: none"> <li>- Keine Operation</li> <li>- Problem im Kopfbereich</li> <li>- Hohe ASA-Klassifizierung</li> <li>- Internistische Behandlung</li> <li>- Keine Probleme an Extremitäten</li> <li>- Keine Schädigung am Bauch</li> <li>- Aufnahme von Außerhalb</li> </ul>
	Bsp.: Patient nach einer Bypass Operation	Bsp.: Patient mit schwerer Lungenentzündung	Bsp.: Schlaganfallpatient

ASA-Klassifizierung: Gebräuchliche Klassifikation zur Einstufung von Patienten hinsichtlich eines Narkoserisikos. Niedrige ASA Werte charakterisieren im allgemeinen gesunde Patienten mit geringem Narkoserisiko, mit steigenden Werten wird von einem steigendem perioperativen Letalitätsrisiko ausgegangen

Sequenzanalysen über den zeitabhängigen Daten wurden auf einzelnen Patientengruppen und auf der gesamten Population ausgeführt. Es konnten nur jeweils Ketten, die einen Zeitraum von maximal 3 Tagen umfassten, betrachtet werden. Regeln, die den Übergang von einem Grad der Funktionseinschränkung eines Organs auf den gleichen Grad an den folgenden Erfassungszeitpunkten beschreiben oder Regeln, die Kombinationen aus verschiedenen Organen, die in ähnliche Kombinationen von Organen mit gleichem Grad der Funktionseinschränkung überführten, traten mit hohem Support und mit hoher Konfidenz auf.

Ein Zeitfenster zur Sequenzanalyse von nur 3 Tagen ist sehr kurz im Vergleich zu Liegedauern von bis zu 90 Tagen. Ein langer Verlauf einer Krankheit ist oft nur von langsamen Änderungen im Krankheitszustand begleitet oder Veränderungen des Zustand der Organsysteme ergeben sich relativ schnell und verharren dann für eine längere Zeit im erreichten Zustand. Die Regeln der Analysen unterstreichen dies. Für die Betrachtungen hinsichtlich der Simulation liefert die Sequenzanalyse in der ausgeführten Form jedoch nicht die gewünschte Art von Information. Hier wären Ergebnisse im Hinblick auf Zustandsänderungen im Zusammenhang mit den zeitlich vorausgehenden Zuständen von Interesse.

## 5. Fazit

Data Mining stellt auch im Bereich der Intensivmedizin ein ernst zunehmendes Werkzeug dar. Die Relevanz von Data Mining Methoden im Bereich der Intensivmedizin wird sich mit wachsender Computerisierung von Intensivstationen und sich etablierenden Datenstandards und damit größer werdenden Fallbasen in Zukunft sicherlich noch vergrößern. Aber auch mit relativ kleinen Fallzahlen können interpretierbare Ergebnisse erzielt werden. Bezüglich der Analyse von Krankheitsverläufen müssen mit anderen Kodierungen der Organzustände und deren Übergänge weitere Möglichkeiten betrachtet werden.

---

## Literatur

- <sup>i</sup> Gaba DM, Fish KJ, Howard SK: Crisis Management in Anesthesiology. New York, Churchill Livingstone, 1994
- <sup>ii</sup> S. Mönk, H.-J. Baldering, J. Vollmer, H. Buggenhagen, W. Heinrichs: „Patientensimulation“, Notfall & Rettungsmedizin, 1999-2:297-306, Springer-Verlag 1999
- <sup>iii</sup> Hersteller des in Mainz benutzten Simulators ist die Firma Medical Education Technologies Inc. (<http://www.meti.com>)
- <sup>iv</sup> Dybowski, Weller, Chang, Gant: „Prediction of outcome in critically ill patients using artificial neural network synthesised by genetic algorithm“ Lancet, 1996 Apr 27, 347 (9009): 1146-50
- <sup>v</sup> T. Weiler, J. Schmidt, H.-J. Baldering, W. Heinrichs: „Qualitätssicherung in der Intensivmedizin: Definition der Inhalte des Kerndatensatzes. Intensivmedizin“, Anästhesiologie & Intensivmedizin 11(39), 575-580, 1998
- <sup>v</sup> T. Weiler, J. Schmidt, H.-J. Baldering, W. Heinrichs: „Entwicklung des Kerndatensatzes: Qualitätssicherung in der Intensivmedizin“, Anästhesiologie & Intensivmedizin 6(39), 1998
- <sup>vi</sup> <http://www.sas.com/offices/europe/germany/solutions/sem.html>