

„Predictive Modeling“

Christian Elsner
David W. Plocher
Stephané Roze
Christian Mazzi

Werkzeug für effizientes Disease Management

Eine Studie untersucht, inwiefern der Ansatz des „Predictive Modeling“, der in den USA bereits zur Versorgung der Patienten genutzt wird, auch auf das deutsche Gesundheitswesen übertragbar ist.

Mit dem methodischen Ansatz des „Predictive Modelings“ managen einige amerikanische Health Management Organizations (HMOs) die Gesundheit ihrer Mitglieder: Über Algorithmen werden potenziell risikoreiche Entwicklungen in Patientengeschichten erkannt, sodass eine Intervention auf der Ebene des Disease Managements oder auch individuell auf der Ebene des Case Managements effektiver eingeleitet werden kann. Geprüft werden muss, ob sich dieser Ansatz auch auf Deutschland übertragen lässt. Struktur und Datensituation scheinen hier auf den ersten Blick ungeeignet zu sein. Die Tatsache jedoch, dass die Systeme nicht komplexe medizinische Datensammlungen erfordern, sondern mit US-Äquivalenten der Daten des § 301 SGB V¹ valide Aussagen treffen können, macht den Ansatz interessant. Zusammen mit US-Anbietern untersucht eine Arbeitsgruppe an der Universität Leipzig, welches Potenzial die Algorithmen – zum Beispiel im Rahmen der Integrierten Versorgung – in Deutschland haben könnten.

Unter proaktivem „Customer Relationship Management“ verstehen Unternehmen, anhand von Kundenprofilen Aussagen über Kundenbedürfnisse treffen zu können, um Kunden gezielt mit Informationen zu versorgen oder Probleme zu erkennen. So könnte den Italienurlauber beispielsweise auch das Buch über Rom interessieren, oder der Kunde mit wiederholtem Druckerpatronenkauf ist an einer Nachfüll-

lösung interessiert. Ein Vergleich mit dem Disease-Management-Ansatz scheint weit hergeholt zu sein: Patientendaten sind komplexer, kommen in der Realität stärker gestreut vor; der Patient wird nicht nur mit Produktinformationen versorgt, sondern ist in einen komplexen Prozess mit verteilten Partnern eingebunden.

Und doch hat man in den USA genau diese Parallele gezogen: Für den US-Markt liegt diese nahe, weil hier neben einer guten Datenlage die HMOs direkt mit der Koordination von Patientenbehandlungen betraut und daher die ideale Schnittstelle für eine solche Koordination sind: Eine Research-Studie der Forrester Gruppe (Veröffentlichung im April 2001) spricht dem Ansatz auch großes Potenzial zu. Danach kann mit diesen Methoden theoretisch die gleiche medizinische Qualität mit der Hälfte des Budgets erreicht werden – für eine Verbesserung der medizinischen Versorgungsqualität mit vorhandenem

Budget stehen demnach enorme Ressourcen zur Verfügung.

Funktionsweise

Der Markt für Algorithmen zur Vorhersage von Krankheitsentwicklungen ist vielfältig. Grundsätzlich lassen sich drei Methoden unterscheiden, mit denen eine Modellierung vorgenommen werden kann (Tabelle). Jeder Ansatz hat seine Stärken und Schwächen, sodass sich keine Strategie klar favorisieren lässt, sondern je nach spezieller Anwendung gewählt werden muss.

• **Empirische Boolsche Regeln:** In dieser Kategorie werden induktive Boolsche Regeln gebildet, die auf Grundlage einer Festsetzung von Grenzwerten für bestimmte Parameter funktionieren: Asthmapatienten könnten zum Beispiel mit einer Regel „... wenn ein Patient mehr als zwei Notfallinterventionen zu einem asthmatischen Anfall innerhalb von sechs Monaten

Tabelle

Methodische Ansätze zum „Predictive Modeling“

Methoden	Vorteile	Nachteile
Empirische Boolsche Regeln („wenn ... dann ... sonst“ – Verknüpfungen)	<ul style="list-style-type: none"> • Klare Modellierung klinischer Zusammenhänge • Strukturierte Anwendung der Regeln 	<ul style="list-style-type: none"> • Inflexibel für statistische Aussagen • Pflegeintensiv und komplex in der „Wartung“
Statistische Techniken (wie Markov-Ketten, Regressionsanalysen, Korrelationsanalysen, lineare Algebra)	<ul style="list-style-type: none"> • Abbilden komplexer und unscharfer Relationen • Scoring/Statistische Aussagekraft möglich 	<ul style="list-style-type: none"> • Pflegeintensiv und komplex in der „Wartung“
Künstliche Intelligenz (mit genetischen Algorithmen oder neuronalen Netzwerken)	<ul style="list-style-type: none"> • Schnelle Generation mit wenig Pflegeaufwand • Schnelle Einarbeitung neuer Daten 	<ul style="list-style-type: none"> • Wenig interpretierbare Regeln als Ergebnis • „Black Box“-Problematik der Anwendung

¹§ 301 SGB V bezeichnet die Grunddaten, die die Krankenhäuser in Deutschland im Rahmen einer Behandlung an Versicherungen übermitteln müssen.

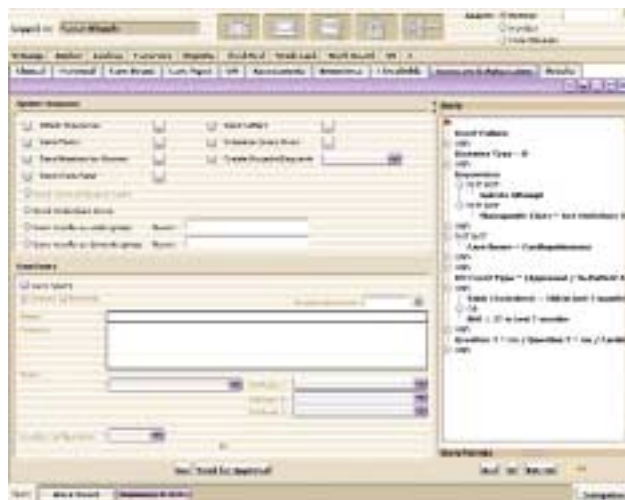
hat, ohne dass er Notfall-Aufnahmen wegen anderweitiger Gründe in den letzten drei Jahren hatte, dann . . ." in einem ersten Schritt eingegrenzt werden. Diese Regeln werden vorwiegend anhand klinischer Erfahrung und aus klinischen Richtlinien gewonnen. Die Anwendbarkeit auf große Personengruppen setzt jedoch ein Werk von weit mehr als 500 bis 1 000 Einzelregeln voraus. Dies macht die Wartung und Validierung eines solchen Ansatzes schwierig und arbeitsintensiv.

• **Statistische Modelle:** Diese identifizieren die Zusammenhänge zwischen klinischen Daten und Entwicklungen. Verwendet werden die in der Tabelle aufgelisteten Methoden. In der Regel werden Daten, die über zwei Jahre erhoben werden, verwendet, um das Modell anhand der Daten eines Jahres zu erstellen und dann mit den Daten des zweiten Jahres zu prüfen. Oftmals werden hier auch verschiedene Studien miteinander verschmolzen und dann mit neuen klinischen Ergebnissen geprüft. Ein Beispiel für ein vielschichtiges statistisches Modell ist der Ansatz von Mellibase (www.mellibase.de). Hier wird ein Modell aus Parametern bei Diabetes verwendet, wie zum Beispiel der aktuelle Nüchternwert und der HbA_{1c} -Wert, um eine Aussage zum Risiko bestimmter Komplikationen zu treffen. Mellibase verwendet die Parameter im Rahmen eines patientenzentrierten „verhaltenstherapeutischen“ Ansatzes, kann Daten aber auch für bestimmte Bevölkerungsgruppen unter gesundheitsökonomischen Aspekten darstellen. Dadurch kann der Einfluss bestimmter Veränderungen, zum Beispiel durch Disease-Management-Programme (DMP), genau abgebildet werden. Die Wartung der statistischen Ansätze ist zwar einfacher als in den Booleschen Regelwerken, aber auch hier ist noch erheblicher Aufwand notwendig. Bezogen auf große Versicherten-

gruppen, gelten außerdem die mit statistischen Modellen verbundenen Nachteile und Fragen: Wurden etwa bei der Erstellung auch alle relevanten Variablen berücksichtigt? Darüber hinaus

können zum Beispiel Kausalitäten nicht klar abgebildet werden: Wird das erhöhte Risiko durch die höhere Zahl behandelnder Ärzte verursacht, oder führt ein erhöhtes Risiko zu einer Behandlung durch mehr Ärzte?

• **Künstliche Intelligenz:** Dieser Ansatz ist die jüngste Herangehensweise im „Predictive Modeling“. Eingesetzt werden hierbei zum Beispiel neuronale Netzwerke. Die Vorgehensweise ähnelt der im vorab beschriebenen Modell: Die Modellierung findet mit Initialstudien- daten statt, die Überprüfung der Daten anschließend mit der – nicht zur Model-



Beispiel für eine Care-Management-Applikation

lierung verwendeten – zweiten Hälfte der Daten aus der Studie. Mit diesem Ansatz wurden zuletzt in den USA sehr gute und teils gegenüber den statistischen Modellen bessere Ergebnisse erzielt. Allerdings haben die Modelle, die ausschließlich auf dieser Technologie basieren, neben dem Vorteil der schnellen Anwendbarkeit einen Nachteil: Sie entstehen meist als „Black Box“ – eine Einsicht in die Verknüpfung und Gewichtung oder eine einfache Veränderung der einzelnen Parameter ist daher meist nicht oder nur sehr schwer möglich.

Ansätze und Ergebnisse

Aktuell kombinieren die US-Anbieter oftmals die verschiedenen Werkzeuge zur Modellierung: So lassen sie die Daten der Krankenversicherten zum Beispiel erst durch eine Reihe von Booleschen Regeln in Untergruppen

aufteilen, um dann mit einer Kombination von statistischen Methoden und neuronalen Netzen Aussagen zu Risiken und Interventionsbedarf zu treffen.

Die Leistungskraft eines exemplarischen „Paketes“ von Regeln wurde in einer retrospektiven Studie mit Daten von 28 900 Versicherten nachgewiesen. In der Studie wurde anhand von zurückliegenden Einjahresdaten eine Projektion zur Entwicklung von Komplikationen vorgenommen und dann mit den vorliegenden Realdaten verglichen. In der Studie wurde eine Komplikation als überdurchschnittliche Kostenentwicklung von mehr als 2 000 Euro Gesundheitsausgaben pro Patient/Jahr definiert. In der durch die Algorithmen identifizierten Untergruppe von Patienten konnte mit den Realdaten eine 3,5fach erhöhte Wahrscheinlichkeit zur Entwicklung einer kostenbeziehungsweise komplikationsintensiven Krankheitsentwicklung gegenüber der Grundpopulation nachgewiesen werden: Während die Komplikationsrate in dieser bei zwölf Prozent lag, war sie in der vorab identifizierten Population mit 39 Prozent weit höher.

Zwar gelang die „Kondensierung“ der Risikopatienten vorerst nur mit einer Sensitivität (= Prozentsatz der Risikopatienten, die als solche identifiziert werden) von zwei Prozent und Werten für den positiven prädiktiven Wert (= Wahrscheinlichkeit, dass ein als Risikopatient klassifizierter tatsächlich Komplikationen entwickelt) von 39,9 Prozent beziehungsweise für den negativen prädiktiven Wert (= Wahrscheinlichkeit, dass ein nicht als Risikopatient klassifizierter tatsächlich keine Komplikationen entwickelt) von 87,8 Prozent.

In aktuellen retrospektiven US-Studien zur Aussagekraft und statistischen Relevanz erweiterter Modelle zeigt sich jedoch, dass noch deutlich höhere Kondensierungen der Risikopopulationen möglich sind. Die führenden Anbieter liegen hier bei Sensitivitäten zwischen 16 und 35 Prozent und ROC-Kurven (= Prozentsatz der gesamten richtigen Aussagen eines Tests zu Risiko beziehungsweise Nicht-Risiko aller geteste-

ten Patienten) zwischen 62 und 82 Prozent. Hier lässt sich einfach kalkulieren, dass es durch die Möglichkeit einer solchen Vorselektion zu einer neuen Effizienz von Disease-Management-beziehungsweise Case-Management-Maßnahmen kommen kann.

Vom Interventionsbedarf zur kontrollierten Intervention

Was nutzt es, vorab zu wissen, welche Patienten ein erhöhtes Risiko besitzen? Medizinisch sinnvoll ist das „Predictive Modeling“ nur als „Puzzleteil“ eines Gesamtprozesses. Bei der Selektion und der Anwendung von Regeln wird daher auf bestimmte Kriterien geachtet, die Grundeigenschaften sicherstellen sollen. Regeln sollen demnach über hohe Sensitivität und Spezifität verfügen. Wichtig sind darüber hinaus die einfache Anwendbarkeit der Regeln, die schnelle Verfügbarkeit der Daten und der direkte Bezug zur klinischen Praxis.

Nur so gelingt es, aus den „Puzzleteilen“ ein sinnvolles Gesamtbild aufzubauen und Programme zu entwickeln, mit denen sich der Ablauf: Interventionsbedarf halbautomatisiert erkennen, geeignete Maßnahmen computerunter-

stützt und medizinisch begründet auswählen und anschließend die Intervention überwachen gestalten lässt. Die US-Anbieter setzen daher zunehmend Lösungen ein, die diese gesamte patientenzentrierte „Wertschöpfungskette“ abdecken können.

Ausgangspunkt ist die Kondensierung von Risikopatienten über das „Predictive Modeling“. Anschließend folgt die halbautomatisierte und mit medizinischem Sachverstand begleitete Assoziation von teilstandardisierten Abläufen mit einzelnen Patienten oder Gruppen von Patienten. Diese Abläufe sind letztlich angepasste Disease-Management-Programme: Von regelmäßigen Patientenaufklärungen über Vorgaben zur Medikamentenversorgung nach aktuellen evidenzbasierten Schemata bis hin zu regelmäßigen Arztbesuchen und Kontrollen bestimmter Blutwerte des Patienten lässt sich alles abbilden (*Abbildung*).

Die Software ermöglicht allen Beteiligten – Krankenhausärzten, niedergelassenen Ärzten, Pflegepersonal, Rehabilitationseinrichtungen, Patienten und deren Angehörigen –, nach den aufgestellten Leitlinien einen Behandlungsablauf zu absolvieren, anzupassen und im Sinne von „Meilensteinanalysen“ zu

überwachen. „Predictive Modeling“ passt sich dabei vollständig in die bei der Integrierten Versorgung angestrebte Wertschöpfungskette zur verbesserten Versorgung ein. Auf Grundlage von um Teildaten ergänzten Äquivalenten des § 301 SGB V steuern derartige Plattformen in den USA bereits mehrere Millionen Versicherte.

Übertragbarkeit

Auch in Deutschland gehen inzwischen Anbieter diesen Weg: So arbeitet beispielsweise die DKV AG, Köln, mit ihrer Sparte „Arztpartner Almeda“ an Modellen, die in diese Richtung zielen. Für ausgewählte DMP werden hier über elektronische Formulare Zusatzdaten erhoben und mit Steuerungsmöglichkeiten verbunden – ein Ansatz, der die computerunterstützte Auswahl und Überwachung von Maßnahmen zulässt. Im Unterschied zu den US-Konzepten fehlen in Deutschland jedoch noch die „Predictive Modeling“-Komponente und der Ansatz, vorhandene Daten (im Sinne der Erhebungen nach § 301 SGB V) zu verwenden. Für die breite Umsetzung und Effektivität der Maßnahmen ist dies ein entscheidender Faktor.

Die universitäre Arbeitsgruppe Medkonsult/Campus Inform untersucht daher mit einer erweiterten Studie, wie die US-Werkzeuge im deutschen Gesundheitswesen implementiert und angewendet werden können (*Grafik*). Zurzeit findet bei vielen Versicherungen eine Konsolidierung der Datenbasis statt, sodass hier ein geeigneter Grundstock entsteht, um mit den Daten zu arbeiten. Die über die Integrierte Versorgung und im Sinne einer Datenkonsolidierung und Datenerhebung um DMP entstehenden Rahmenbedingungen können daher künftig die Voraussetzungen für eine Kombination mit den US-Ansätzen bieten.

■ Zitierweise dieses Beitrags:
Dtsch Arztebl 2004; 101: A 1724–1726 [Heft 24]

Anschrift für die Verfasser:
Dr. med. Christian Elsner
Part-Time MBA Programm
Handelshochschule Leipzig
Jahnallee 59, 04109 Leipzig
E-Mail: christian.elsner@mba.hhl.de

Grafik

Ineinandergreifen der Systeme – Ansatzpunkte

