

## PIPES<sub>med</sub>: Ein flexibles Werkzeug zur Verarbeitung kontinuierlicher Datenströme in der Medizin

Seeger B<sup>2</sup>, Krämer J<sup>2</sup>, Penzel T<sup>3</sup>, Lenz R<sup>1</sup>

<sup>1</sup>Institut für Medizinische Informatik, Universitätsklinikum Gießen und Marburg GmbH, Standort Marburg

<sup>2</sup>Lehrstuhl für Datenbanksysteme, Universität Marburg

<sup>3</sup>Schlafmedizinisches Labor, Universitätsklinikum Gießen und Marburg GmbH, Standort Marburg  
lenzr@staff.uni-marburg.de

**Einleitung und Fragestellung** Die fortschreitende Entwicklung von elektronischen Geräten und insbesondere Sensoren für das Monitoring von Patienten führt zukünftig zu deren verstärktem Einsatz in nahezu allen Bereichen der Medizin, wie beispielsweise in der Intensivmedizin, im Schlaflabor und in zunehmendem Maße auch bei der Überwachung chronisch kranker Patienten. Dabei ergibt sich bereits heute die Notwendigkeit nicht immer nur größere Datenmengen zu verwalten, sondern vor allem auch die resultierenden, kontinuierlichen Datenströme zeitnah zu analysieren. In Hinblick auf den Umgang der in den Anwendungen auftretenden massiven Datenvolumina lassen sich derzeit prinzipiell zwei Verfahrensweisen unterscheiden.

1) Im Rahmen des Patienten-Monitorings werden derzeit zahlreiche heterogene Datenströme nach dem Prinzip „Anzeigen und Vergessen“ erzeugt. Alarme werden meist nur auf der Basis einzelner Datenströme generiert, wenn die beobachteten Daten bestimmte Schwellwerte überschreiten. In vielen Fällen wird eine Risikosituation aber nicht durch einen überschrittenen Schwellwert, sondern eher durch eine hohe Änderungsrate im Datenstrom oder multivariate Eigenschaften aus verschiedenen Datenströmen indiziert. Änderungsraten und multivariate Eigenschaften werden jedoch derzeit nur unzureichend berücksichtigt, so dass solch komplexere Gefahrensituationen nur durch eine medizinische Fachkraft entdeckt werden können. Um dennoch automatisch eine höhere Sensitivität zu erreichen, werden daher oft niedrige Schwellwerte angesetzt, was zu vielen falsch positiven Alarmen führt. Dies birgt die Gefahr einer Desensibilisierung der Mitarbeiter und erhöht das Fehlrisiko. Die Möglichkeit zur Verbesserung des Patienten-Monitorings durch die Berücksichtigung einer geeigneten Fusion von Datenströmen liegt auf der Hand (vgl. auch [1]).

2) Die erfassten Daten werden im Originalzustand zunächst vollständig zur späteren, detaillierten Analyse abgespeichert und archiviert. Mit der Zeit entstehen hierdurch riesige Datenarchive, deren vollständige Analyse auch mit modernsten Computern äußerst zeitaufwändig ist. Um diesem Problem entgegenzuwirken, wäre es besser, statt den Originaldaten nur die relevanten bzw. interessanten Kennzahlen und Charakteristiken aus den Datenströmen des Monitorings zu extrahieren, geeignet zu transformieren und diese dann abzuspeichern. So könnte es vorteilhaft sein, wenn man bereits während des Monitorings eine Fusion verschiedener Datenströme mit einer anschließenden Analyse vornehmen würde, um auf diese Weise später ohne aufwändige Berechnungen direkt auf relevante multivariate Datencharakteristiken und Zustandsindikatoren zugreifen zu können.

**Material und Methoden** PIPES (A Public Infrastructure for Processing and Exploring Streams) ist ein Werkzeug zur flexiblen Spezifikation von Operatoren auf Datenströmen, das an der Universität Marburg entwickelt wurde [2]. Mit Hilfe von PIPES ist es möglich, Funktionalität zur Anfrageverarbeitung wie sie derzeit in Datenbanken genutzt wird, auf die Datenstromverarbeitung zu übertragen. Ein entscheidender Vorteil ist, dass die Daten mittels PIPES direkt beim Eintreffen verarbeitet werden, d.h., man muss diese nicht zuerst in einer Datenbank ablegen. Anfragen in PIPES werden kontinuierlich ausgewertet, d.h., sobald neue Daten eintreffen, werden die Ergebnisse unverzüglich berechnet und bereitgestellt. Die Infrastruktur PIPES bietet durch ihre reichhaltige, in Java implementierte Funktionalität vielseitige Anbindungsmöglichkeiten an externe Datenquellen und Informationssysteme. Der Einsatz von PIPES in der medizinischen Informatik ist eine weitere Anwendung der in einem DFG-Projekt entwickelten allgemeinen Operatoralgebra für die Datenstromverarbeitung, die bereits erfolgreich in anderen Applikationen, wie beispielsweise der Analyse von Sensordaten in der Fabrikautomation angewendet wurde [4]. PIPES ist für die obigen medizinischen Fragestellungen interessant, da die Operatoralgebra unter anderem Operatoren beinhaltet, die eine Fusion von Datenströmen erlauben [3]. Auf diese Weise ist es möglich Werte bzw. Signale aus verschiedenen Datenströmen miteinander in Verbindung zu setzen. Eine solche Fusion kann dabei sehr flexibel definiert werden. Allerdings erfordert die Anpassung von PIPES auf den medizinischen Kontext eine enge Zusammenarbeit zwischen Medizinern und Informatikern, da etwa die Rahmenbedingungen für die Fusion von medizinischen Parametern und Erfahrungen abhängen, die durch die Informatiker formal erfasst und in Verbundprädikate überführt werden müssen. Zudem bietet PIPES die Möglichkeit, Anfragen auf zeitlich über den Daten gleitenden Fenstern zu berechnen, d.h., dass die Auswertung einer Anfrage auf ein Zeitfenster bezogen wird, wie z.B. die letzten 30 Minuten, oder aber auf ein Fenster fester Größe, wie etwa die letzten 500 Messwerte eines Signals. Ein weiterer Bestandteil von PIPES beschäftigt sich mit der kontinuierlichen Analyse von Datenströmen. Es ermöglicht die Erstellung anwendungsspezifischer Statistiken über den bisher konsumierten Teil eines Datenstromes, welche wiederum zur Generierung von Prognosen und Abschätzungen über den in der Zukunft liegenden Datenstromteil verwendet werden können. Ein einfaches Beispiel hierfür ist die kontinuierliche Berechnung der Varianz eines Eingangssignals. Die Analysefunktionalität von PIPES umfasst auch die Erkennung von Änderungen in der Charakteristik eines Datenstromes. Generell werden hierzu Methoden aus der Mathematik und dem Bereich des Data Mining auf Datenströme angewendet.

Zur Simulation von kontinuierlichen medizinischen Messdaten wurde umfangreiches Datenmaterial aus dem Schlafmedizinischen Labor herangezogen, das bislang im Rahmen von Langzeitmessungen aufgezeichnet und nachträglich ausgewertet wurde. Um den Nutzen durch PIPES in einem solchen Szenario zu evaluieren, musste zunächst eine Anbindung hergestellt werden. Ein spezieller Konverter transformiert die medizinischen Daten direkt in das interne Format von PIPES. Um möglichst praxisnahe Ergebnisse zu erzielen, wurden die aufgezeichneten Signale in Echtzeit abgespielt, d.h., mit Originalverzögerungen und Abstraten.

**Ergebnisse** Die Anbindung an PIPES und die Simulation auf Basis der aufgezeichneten medizinischen Datenströme konnte erfolgreich und mit relativ geringem Aufwand umgesetzt werden, da einerseits für die Daten ein vorgegebenes Format existiert und PIPES andererseits dedizierte Mechanismen für den Import und Export von Datenströmen bereitstellt. Durch die Simulationen konnte gezeigt werden, dass bereits mit der Standardfunktionalität zur Anfrageverarbeitung in PIPES, die im Wesentlichen der von Datenbanksystemen entspricht, einfach und in kurzer Zeit kontinuierliche Anfragen über medizinischen Daten formuliert werden können, die zum Beispiel komplexe Alarmbedingungen über mehreren Datenströmen umsetzen. Da PIPES in der Lage ist, viele kontinuierliche Anfragen über einer Menge von Datenströmen zur gleichen Zeit effizient auszuwerten, lassen sich eine Vielzahl unterschiedlichster Alarmbedingungen in nahezu Echtzeit überprüfen und gleichzeitig wichtige Kenndaten berechnen sowie anzeigen.

Ein weiterer Vorteil, der sich durch den Einsatz von PIPES ergibt, besteht darin, dass die Daten ohne eine Zwischenspeicherung direkt bei deren Eintreffen verarbeitet werden und unmittelbar die Anfrageergebnisse bereitgestellt werden. Falls eine Aufzeichnung relevanter Daten erwünscht ist, besteht die Möglichkeit die Ausgabeströme in Datenbanken abzuspeichern oder an externe Informationssysteme weiterzuleiten. An dieser Stelle sei erwähnt, dass durch die Verarbeitung mit PIPES die Datenvolumina der Ausgabeströme um ein Vielfaches kleiner sind als die ursprünglichen Eingabeströme. Dies ist darauf zurückzuführen, dass irrelevante Daten durch die kontinuierlichen Anfragen frühzeitig herausgefiltert wurden. Typische Operatoren, die bei diesen Anfragen verwendet wurden, waren Filterungs-, Fusions- und Verdichtungs-, und Analyseoperatoren. Diese Operatoren weisen die inhärente Eigenschaft auf, dass deren Ausgabestrom im Allgemeinen signifikant kleinere Datenvolumina haben als die Eingabeströme. Durch die geeignete Definition kontinuierlicher Anfragen lassen sich folglich Ausgabeströme erzeugen, die lediglich auswertungsrelevante Daten umfassen, für die eine Archivierung lohnenswert erscheint.

**Diskussion** Die bisherigen Ergebnisse haben gezeigt, dass der Einsatz von PIPES zur Verarbeitung medizinischer Datenströme nicht nur die effiziente und zeitnahe Verwaltung großer Datenmengen ermöglicht, sondern auch ein hohes Potential zur Verbesserung von Monitoring-Prozessen und zur Minimierung von Fehlern beinhaltet. Geplant ist die Ausweitung des Einsatzes durch Anwendungen in der Intensivmedizin, sowie die Spezifikation eines Werkzeugkastens mit typischen wiederverwendbaren Operatoren, die zur Analyse medizinischer Datenströme flexibel eingesetzt und kombiniert werden können. Die medizinspezifischen Erweiterungen von PIPES, wie beispielsweise die Konverter und Operatorkonfigurationen für bestimmte medizinische Anwendungsgebiete, sollen unter dem Namen PIPES<sub>med</sub> öffentlich zur Verfügung gestellt werden.

## Literatur

- [1] Burkom HS, Murphy S, Coberly J, Hurt-Mullen K. Public health monitoring tools for multiple data streams. MMWR Morb Mortal Wkly Rep 2005; 54 Suppl: 55-62.: 55-62.
- [2] Krämer J, Seeger B. PIPES – A Public Infrastructure for Processing and Exploring Data Streams. ACM SIGMOD Conference 2004; 925-926.
- [3] Cammert M, Heinz C, Krämer J, Seeger B. Anfrageverarbeitung auf Datenströmen. Datenbankspektrum 2004; Vol. 11, 5-13.
- [4] Cammert M, Heinz C, Krämer J, Riemenschneider T, Schwarzkopf M, Seeger B, Zeiss A. Stream Processing in Production-To-Business Software. IEEE ICDE 2006.